Development of a novel high throughput method for identifying phage-host pairs in an extreme environment

Israel Temiloluwa Olonade

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Department of Biotechnology,

University of the Western Cape,

Bellville, Cape Town, South Africa.

Supervisor: Professor Marla Trindade (Tuffin)

August, 2016

Declaration

I, Israel Temiloluwa Olonade, hereby declare that "**Development of a novel high throughput method for identifying phage-host pairs in an extreme environment**" is my own work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Date:……

Signed:…….

Abstract

There are approximately 10^{31} bacteriophages in the biosphere, outnumbering bacteria 10:1, hence, the dynamic and specific nature of phage-host interactions exerts significant influence on microbial communities. Bacteriophages also represent the reservoir of the highest known genetic diversity making them a potential source of novel biotechnological products. However, the isolation of novel bacteriophages is limited by the observation that less than 1% of bacterial hosts have been cultured.

This study aimed to bypass this problem by developing novel culture independent approaches to improve our ability to isolate novel phage-host pairs. Samples were collected from an abandoned copper prospecting site near the Gobabeb Desert Research and Training Station and a Salt lake located in the Swakopmund region of the Namibian desert. Two approaches were explored in this study namely viral tagging and reverse metaviromics.

For viral tagging, fluorescently labelling the environmental phage fraction before challenging the environmental bacterial fraction with tagged phages proved difficult. This was most likely due to the complex interaction of the labelling agent with phages and requires further studies.

For the reverse metaviromics approach, total DNA from the environmental phage fractions was extracted, sequenced and analyzed for novel phages. Analysis of the phage diversity showed that the copper site was dominated by tailed viruses as has been shown for other extreme arid environments. However, the saline site was atypical of marine environments, with tailed viruses being the most abundant, suggesting that the diversity present is not only driven by salinity.

Using the metaviromic sequence data to guide the selection of potential bacterial hosts, two strategies were employed. In the first, putative hosts were predicted based on similarity of

phage sequences to those identified in databases. Media targeting these specific genera were employed, 8 bacterial species were isolated and based on 16S rRNA similarity to the closest known species were identified as *Halomonas caseinilytica, Halomonas eurihalina*, *Halomonas sinaiensis*, *Idiomarina loihiensis*, *Marinobacter xestospongiae*, *Virgibacillus salarius* and two *Salinivibrio* species. The 16S rRNA analysis also suggested that *H. sinaiensis*, *V. salarius* and both *Salinivibrio* species are novel. All 8 isolates were challenged with the environmental phage fraction. A novel phage, SMHB1, was isolated on one of the *Salinivibrio* spp. and is only the second characterized phage ever described for this genus. SMHB1 is a 32 kb myovirus, with a head diameter of 56 nm, and a tail length of 106 nm.

The second approach involved the design of fluorescently labelled probes targeting phages identified from the metaviromic sequence data. In a control *E. coli* system to detect cloned phage DNA fragments, 87% of the interrogated cells showed significant hybridization of the phage specific probe to the target. The optimized method was applied to a simulated environmental bacterial fraction and a detection limit of 1:100 was observed for the bacteria containing the phage DNA fragment of interest.

This study demonstrates the possibility of improving the specificity of isolating phage-host pairs in a culture-independent manner by incorporating sequence data in the experimental design; and contributes to our knowledge of the phage diversity of an understudied extreme environment.

Keywords: extreme environments, metaviromics, viral diversity, phage-host interactions, viral tagging, fluorescence *in situ* hybridization.

Acknowledgments

My profound gratitude to:

- God. I would not have made it without You. No brainer. I owe You everything. I am so grateful.
- My Supervisor, Professor Marla Trindade. Thank you for the opportunity, guidance, and for providing an excellent research environment.
- Lonnie van Zyl, my “Co-supervisor”, for the advice, lectures, corrections, suggestions, reagent chasing and arguments. Thank you for everything.
- My Dad. A great example of love and sacrifice. I hope I can repay you.
- My Mum. For the umpteenth time, yes, I have eaten, but no, I am not telling you what I ate. Love you mum.
- Emmanuel and Christianah. Wonderful siblings and the reason I can’t rest on my oars.
- Seun Akinyemi #ThatIsAll.
- IMBMers. You have been wonderful in every way.

Dedication

To Mrs Akinyemi. You left way too early but gave me the greatest gift of all.

Table of Contents

Table of Figures

List of Tables

Abbreviations

°C	Celsius
µg	Microgram
µg/ml	Microgram per millilitre
µl	Microlitre
µm	Micrometre
µM	Micromolar
ATP	Adenosine triphosphate
BLAST	Basic local alignment tool
BLASTn	Nucleotide BLAST
BLASTp	Protein BLAST
BLASTx	Protein databases using a translated nucleotide query
bp	Base pair
BSA	Bovine serum albumin
$CaCl_2$. $6H_20$	Calcium chloride hexahydrate
CI	Chloroform isoamyl alcohol
Contig	Contiguous
CTAB	Cetyltrimethylammonium bromide
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphates
EDTA	Ethylenediaminetetraacetic acid
FACS	Fluorescence activated cell sorting
$FeCl_3$	Iron (III) chloride

$Fe(NH_4)_2(SO_4)_2.\ 6H_2O$	Ammonium iron(II) sulfate hexahydrate
FISH	Fluorescence *in situ* hybridization
FSC	Forward scatter
HCl	Hydrochloric acid
IMBM	Institute for Microbial Biotechnology and Metagenomics
Kb	Kilobase pairs
KBr	Potassium bromide
KCl	Potassium chloride
K_2HPO_4	Potassium phosphate dibasic
K_2S0_4	Potassium sulphate
LB	Luria-Bertani
M	Molar
Mbp	Megabase pair
$MgCl_2.\ 6H_20$	Magnesium chloride hexahydrate
$MgSO_4.\ 7H_20$	Magnesium sulphate heptahydrate
Min	Minute(s)
ml	Millilitre
mM	Millimolar
NaBr	Sodium bromide
NaCl	Sodium chloride
$NaHCO_3$	Sodium bicarbonate
NCBI	National Centre for Biotechnology Information
ng	Nanogram
nm	Nanometres
OD	Optical density

ORF	Open reading frame
PBS	Phosphate buffered saline
PCI	Phenol/Chloroform/Isoamyl alcohol
PCR	Polymerase chain reaction
RNA	Ribonucleic acid
rpm	Revolutions per minute
rRNA	Ribosomal ribonucleic acid
Sec	Seconds
SSC	Saline sodium citrate
TAE	Tris-Acetate-EDTA buffer
tBLASTx	translated nucleotide databases search using a translated nucleotide query
TE	Tris-EDTA buffer
TEM	Transmission electron microscopy
U	Unit
UV	Ultraviolet
Volts/cm	Volts per centimetre
w/v	Weight per volume

1

LITERATURE REVIEW

1.1 DEFINITION

Bacteriophages (informally, *phages,* from the Greek word '*phagein'* meaning "to eat") are viruses that infect bacteria (Clokie *et al*., 2011; Ghannad & Mohammadi, 2012). Typically, phages consist of genetic material (DNA or RNA) enclosed within a protein coat (Hankin, 1896; Prescott, 1993). With an estimate of about 10^{31} viral particles present in the environment, they are known to be widely distributed in nature, thought to outnumber bacteria 10:1 and play an important role in shaping microbial diversity and ecology (Fuhrman, 1999; Clokie *et al*., 2011; Ghannad & Mohammadi, 2012).

1.2 HISTORY OF PHAGES

There has been a fair amount of debate over the discovery of bacteriophages. The first recorded mention of phage-like activity was in 1896 when British Bacteriologist, Ernest Hankin, observed antibacterial activity against *Vibrio cholerae* in the water from the Ganges and Jumna rivers in India. He suggested that the spread of the Cholera epidemic was limited by an unidentified substance which was heat labile and passed through fine porcelain filters. A similar occurrence was reported in 1898 by Russian Scientist, Gamaleya, while investigating *Bacillus subtilis* (Samsygina & Boni, 1984). However, phages were not further explored until 1915 when Fredrick Twort reported a similar phenomenon and hypothesized that the unknown antibacterial agent might be a virus.

Bacteriophages were 'officially' discovered in 1917 by Felix d'Herelle, a French-Canadian microbiologist at the Institut Pasteur in Paris. He had also observed the phage phenomenon in 1910 in Mexico while investigating biological methods for controlling an epizootic event in locusts. During the summer of 1915, d'Herelle was commissioned to investigate the outbreak of severe heamorrhagic dysentery among French troops at Maisons-Laffitte, on the outskirts of

Paris. His earlier ‘phage’ observation formed the basis of his hypothesis for producing an effective vaccine against dysentery. He made bacterium-free filtrates of patient’s fecal matter and mixed them with *Shigella* strains isolated from the patients. Experimental animals were inoculated with a portion of the mixture and the rest was spread on agar plates and observed. Small, clear zones were first observed on those plates and were initially called *taches* (plaques), then *taches vierges* (clear plaques) (d’Herelle, 1917).

d'Herelle first isolated phages in 1916 and presented his findings at the meeting of the Academy of Science in September 1917. He also proposed the name ‘bacteriophage’. His findings were accepted and published during meeting proceedings (d’Herelle, 1917). His publication caused a dispute in the scientific community as some argued that Twort discovered phages first (Twort, 1915). d'Herelle continued his research to prove that bacteriophages where viruses and not just enzymes as a few fellow researchers suspected (Bordet & Ciuca, 1921). The argument about discovery eventually subsided, phages were accepted as independent biological entities and the associated antibacterial activity was referred to as the “Twort-d'Herelle phenomenon" and, later, the "bacteriophage phenomenon” (Sulakvelidze *et al*., 2001).

In the following decades, bacteriophages played significant roles in some of the major advancements in our understanding of biology. T2 and T4 phages were used in the experiments to determine that genetic information was carried by DNA rather than protein (Hershey, 1952). The first successful gene mapping used the T4 phage genome (Benzer, 1955). T4 was also used to demonstrate that DNA undergoes discontinuous replication (Okazaki *et al*., 1968). Lambda phage was used in experiments during which restriction endonuclease modification systems were first detected (Bertani & Weigle, 1953), and also used for understanding gene regulation (Ptashne *et al*., 1980), as well as a vector for other molecular biology applications (Sambrook *et al*., 1989).

Recently, phages are being studied as an important factor in shaping microbial diversity (Thurber, 2009), as potential sources of novel enzymes (Hughes *et al*., 1998; Fischetti, 2010; Nelson *et al*., 2012), and as model systems for gene delivery and expression (Larocca *et al*., 2002). Improvements in sequencing technologies have also led to an unprecedented access to bacteriophage genome sequences which are expected to further increase our understanding of phage genomics and biology.

1.3 PHAGE CLASSIFICATION

The International Committee on Taxonomy of Viruses (ICTV) classifies viruses based on morphology and nucleic acid composition among other criteria.

Caudovirales (tailed bacteriophages) account for over 96% of currently known phages and perhaps the majority of phages in nature (Ackermann, 2007). A second Order, *Ligamenvirales*, consisting of two rod shaped phages was recently described (Prangishvili & Krupovic, 2012). Other known phages are currently unassigned to any Order. No phages have been described for some bacteria phyla such as *Aquificae* and *Chrysiogenes* largely because their presence has not been investigated (Ackermann, 2007; Ackermann, 2009). Table 1.1 below summarizes known phage families and their characteristics:

Table 1.1: An overview of phage families. Adapted from Ackermann 2007; Prangishvili & Krupovic 2012 and modified.

Order	Family or Group	Genus	Type member	Particle Morphology
Caudovirales	*Myoviridae*	*T4-like viruses*	*Enterobacteria* phage T4	Icosahedral head, contractile tailed
		P1-like viruses	*Enterobacteria* phage P1	
		P2-like viruses	*Enterobacteria* phage P2	
		Mu-like viruses	*Enterobacteria* phage *Mu*	
		SP01	*Bacillus* phage SP01	
	Podoviridae	*T7-like viruses*	*Enterobacteria* phage T7	Icosahedral head, short tailed
		M29-like viruses	*Bacillus* phage M29	
		P22-like viruses	*Enterobacteria* phage P22	
	Siphoviridae	*8-like viruses*	*Enterobacteria* phage 8	Icosahedral head, non- contractile tailed
		T1-like viruses	*Enterobacteria* phage T1	
		T5-like viruses	*Enterobacteria* phage T5	
		c2-like viruses	*Lactococcus* phage c2	
		L5	*Mycobacterium* phage	

Ligamenvirales	*Lipothrixviridae*	*Alphalipothrixvirus*, *Betalipothrixvirus*, *Gammalipothrixvirus* *Deltalipothrixvirus*	*Acidianus* filamentous virus 1	Rod
	Rudiviridae	*Rudivirus*	*Sulfolobus islandicus* rod-shaped	Rod
Unassigned	*Cystoviridae*	*Cystovirus*	Ø6	Isometric
	Inoviridae	*Inovirus*	*Enterobacteria* phage M13	Rod
		Plectrovirus	*Acholeplasma* phage MV-L51	
	Leviviridae	*Levivirus*	*Enterobacteria* phage MS2	Icosahedral
		Allolevirus	*Enterobacteria* phage Q$	
	Microviridae	*Microvirus*	*Enterobacteria* phage ØX174	Icosahedral
		Spirovirus	*Spiroplasma* phage 4	
		Bdellomicrovirus	*Bdellomicrovirus* phage MAC1	
		Chlamydiamicrovirus	*Chlamydia* phage 1	
	Plasmaviridae	*Plasmavirus*	*Acholeplasma* phage L2	pleiomorphic
	Tectiviridae	*Tectivirus*	*Enterobacteria* phage PRD1	Icosahedral
	Corticoviridae	*Corticovirus*	*Alteromonas* phage PM2	Icosahedral

While the composition of nucleic acids contained within bacteriophages has been observed to differ, ranging from ssRNA to ssDNA to dsDNA, the majority contain double stranded DNA within the capsid (Ackermann, 2007). Phages can also be classified based on the presence or absence of an envelope around the genetic material. The envelope is a lipid bilayer surrounding the capsid and is composed of some bacterial host proteins, phospholipids and some viral glycoproteins that aid host infection (Tolonen *et al*., 2001; Nisole & Saïb, 2004). Plasmaviruses possess such envelopes. Figure 1.1 below shows the classification of phages based on nucleic acid composition.

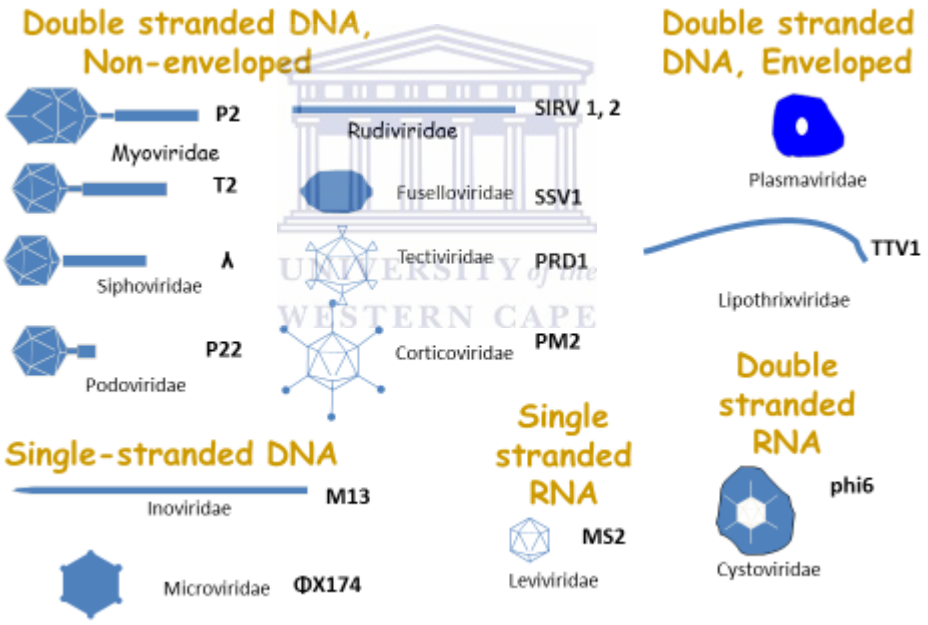

Figure 1.1: Classification of phages based on nucleic acid composition (http://agridr.in) (Accessed: 15-10-2014).

1.4 GLOBAL PHAGE DIVERSITY

While the discovery of phages and application of phage model systems has contributed greatly to molecular biology, not until recently has the impact of viruses on ecosystems been fully appreciated (Bogovazova *et al*., 1991; Cairns *et al*., 2007). Bacteriophages are ubiquitous,

numbering from about 10^7 to 10^9 per gram of top soil or sediment (Danovaro & Serresi, 2000; Hewson *et al*., 2001) to 10^{10} per liter of surface sea water (Bergh *et al*., 1989). Phages are now recognized to contribute significantly to nutrient sink in the ocean by being major predators of bacteria and archaea (Wilhelm & Suttle, 1999).

Phages have also been recognized as major contributors to genetic exchange in the environment and are estimated to transduce 10^{25} to 10^{28} base pairs of DNA per year in the marine environment alone (Paul, 1999).

There are an estimated 10^{31} bacteriophages in the biosphere, outnumbering bacteria 10 to 1 (Fuhrman, 1999; Wommack & Colwell, 2000; Ashelford *et al*., 2003), hence, the dynamic and specific nature of phage-host interactions exerts significant influence on microbial communities (Hambly & Suttle, 2005; Clokie *et al*., 2011). For example, cyanobacteria of the genus *Synechococcus* are among the most abundant picophytoplankton in the ocean, accounting for up 30 % of total primary production (Waterbury *et al*., 1986). The concentration of *Synechococcus* specific phages have been found to range between 10^2 to 10^5 per ml in near shore and offshore waters (Waterbury & Valois, 1993; Suttle & Chan, 1994). It was reported that up to 3% of *Synechococcus* species isolated from various marine environments contain mature phages and these phages may be responsible for up to 14% of cyanobacterial mortality on a daily basis (Suttle & Chan, 1994). This provides an indication of the correlation between phage persistence in the environment and the turnover of the susceptible bacterial population (Wilhelm & Suttle, 1999).

Haloalkaline environments are thought to have the highest abundance and diversity of viruses of all types of environments, with the highest viral count documented to be 2 x 10^9 per ml in the hypersaline Mono Lake, California (Brum *et al*., 2005). The impact of such exceptionally high abundance on trophic levels was recently demonstrated in the East African Rift Valley

lakes where a short food chain involving cyanophages, the cyanobacterium *Arthrospira fusiformis* and the Lesser Flamingo exists (Peduzzi *et al*., 2014). The cyanobacterium is the most important source of food for the flamingo which has recently been classified as 'near-threatened' (IUCN Red List, 2016: http://www.iucnredlist.org). However, the population of *A. fusiformis* collapses drastically and unpredictably and this fluctuation has been reported as the major factor influencing the population and distribution of the Lesser Flamingo (Vareschi & Jacobs, 1985; Krienitz & Kotut, 2010). The study showed that cyanophage infection was directly responsible for the substantial lysis of *A. fusiformis*, hence the drastic reduction in flamingo population. The study also made an interesting observation that although the virus to bacteria ratio constantly fluctuated, the difference remained in the same order of magnitude. This suggests that viruses were actively involved in keeping bacterial populations at a threshold.

The study of phage diversity is complex and usually incomplete. Many studies evaluate only the presence of free phage particles while neglecting the impact of prophages. It is known that most bacteria contain more than one prophage within their genomes and it has also been reported that approximately 60% of bacteria in the marine environment contain inducible prophages (Cochran & Paul, 1998). However, these investigations generally fail to account for the potential increase in the number of phage particles as a result of an induction event. The relatively low incidence of lysogeny may be attributed to fluctuations in temperature and nutrients (Jiang & Paul, 1994; Wilson *et al*., 1996; Cochran & Paul, 1998). It is therefore unsurprising that high concentrations of phage particles have been reported in coastal water and oceans, and even higher concentration in lakes depending on season and geographical location (Kirby *et al*., 1994; Wommack & Colwell, 2000). A study of phages in Chesapeake Bay, an eutrophic estuarine water body, found bacterial density to be 10^6 cells per ml and phage concentration to be 10^7 particles per ml with variations dependent on the season (Wommack *et*

al., 1999). There were about 5 to 10 viral particles per bacterial cell per sampling site. The study also hypothesized that the community contained about 10 to 50 different bacterial species and 100-300 different phage strains. This resulted in the suggestion that phages are the most abundant biological entities in the biosphere (Riley, 2006).

1.5 PHAGE GENOMICS

Phage ΦX174, a 5375 bp single stranded DNA phage, was the first whole genome to be sequenced largely due to the relatively small size and ease of isolation (Sanger *et al*., 1977). This was followed by the whole genome sequencing of the double stranded DNA phage lambda phage at 48502 bp (Sanger *et al*., 1982). Advancements in DNA sequencing technologies have seen the number of complete phage genomes increase exponentially. Currently, there are 1452 complete phage genomes on the NCBI database. The increased interest in sequencing phages can be attributed to two major reasons. First, the realization that phages are more abundant than initially thought and that the degree of relatedness of known genomes is such that insight can be gained into the evolutionary mechanisms that shape such abundance. Second, the increasing application of phages and derived products in developing industrial, biotechnological, and clinical tools (Hatfull, 2008).

Phage genomes range in size from 2435 bp (*Leuconostoc* phage L5) to 497513 bp (*Bacillus* phage G) (Edwards & Rohwer, 2005; Deschavanne *et al*., 2010). However, phage genome size distribution is not homogenous, possibly due to biases in current isolation techniques which include size exclusion using filters and propagation only on bacterial hosts that can be cultured (Serwer *et al*., 2007). Hence, the size distribution observed from sequences may not necessarily reflect the existing distribution in nature. Also, the small number of sequenced phages with large genomes can be partially attributed to the observation that bacteriophages with large

heads, hence big genomes, typically form very small plaques on agar plates and can be easily overlooked during isolation (Serwer *et al*., 2007).

The majority of sequenced phages are double stranded DNA-containing siphoviruses ranging between 30 to 60 kb which account for about 55% of all known phages (Deschavanne *et al*., 2010). This reflects the observation by electron microscopy that tailed phages most likely dominate the biosphere (Ackermann, 2007) and supports the finding that genes responsible for virion structure and assembly typically account for at least 15 kb of the whole genome, with the tape measure protein gene of siphoviruses alone accounting for between 1.5kb to 6 kb (Hatfull, 2008).

Comparative analysis of phage genomes has revealed interesting themes. For example, most siphoviruses have syntenic regions among genes that code for structural and assembly proteins (Casjens, 2005). The head and tail genes are usually arranged together with the head genes situated 5' of the tail genes (Figure 1.2). The head genes usually contain a protease, the portal protein, a scaffold protein, one or two terminase subunits, and the major capsid subunit while the tail genes contain the major tail subunit, the tape measure protein and other minor tail proteins (Casjens, 2005). In some phages, such as MU and λ, the tail genes contain two overlapping open reading frames which are expressed through a programmed translational frameshift (Xu *et al*., 2004). The length of the tape measure protein gene is directly proportional to the tail length of the phage (Katsura & Hendrix, 1984). This syntenic organization is conserved even in phages that have no similarity both at a nucleotide level and across predicted proteins (Casjens *et al*., 1992). This is therefore thought to be an early evolutionary feature.

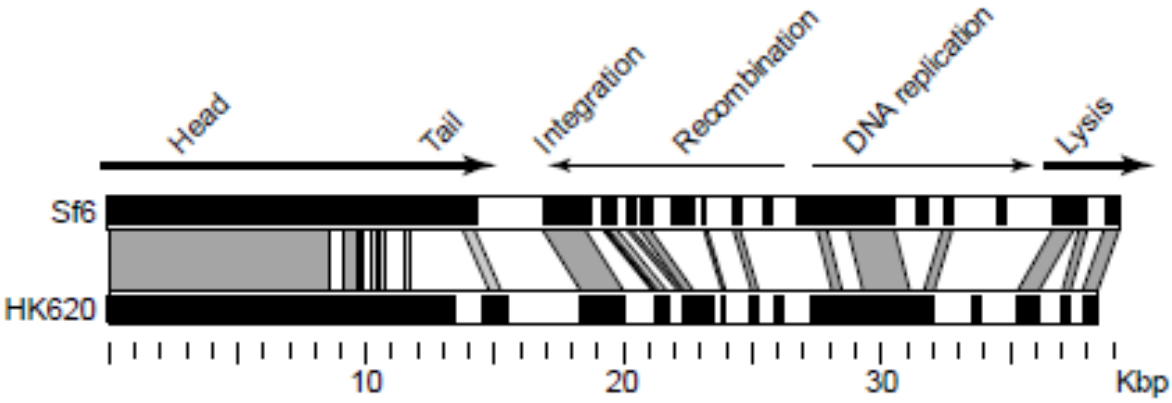

Figure 1.2: Qualitative mosaic relationship between the whole genomes of phage Sf6 and HK620, with their circular genomes opened at the 50 kbp -end of their small terminase genes. Above, the arrows denote the major transcripts (thin arrows represent early operons; thick arrows represent the late operon), and representative functions of both genomes are provided. Black horizontal bars indicate regions of sequence similarity, and grey areas between the genomes indicate regions of >95% (dark), 90–95% (medium) and 80–90% (light) nucleotide sequence identity (Casjens, 2005).

A few variations of this syntenic theme have been observed. For example, the lysis genes in mycobacteriophages are found in different locations with respect to the structural and assembly genes. In some mycobacteriophages such as TM4, the lysis genes are usually found downstream of the minor tail protein genes (Ford *et al*., 1998) while in others like phage L5, the lysis genes are found upstream of the terminase gene usually between the terminase and the *cos* site (Hatfull & Sarkis, 1993).

Also, many mycobacteriophages have an integration cassette which is usually located close to the centre regardless of the size of the genome. However, one instance was observed where the integration cassette was found within the tail genes. This was thought to be due to the presence of a secondary attachment site within the genome (Morris *et al*., 2008).

This syntenic organization is less defined in larger phage genomes (>125 kb) and is generally limited to structural genes where most of the conserved regions are punctuated by hyper plastic regions (Casjens, 2005; Comeau *et al*., 2007). These hyper plastic regions contain relatively

small genes whose functions are mostly unknown but are suspected to be involved in phage adsorption to the bacterial host and protection of the genetic content of the phage from modification by the host (Rifat *et al*., 2008).

The presence of extra genes with no known function is a common feature of phage genomes. These genes are referred to as 'morons' and are characteristically inserted between two predicted phage genes. Morons are thought to be residues of horizontal gene transfer (Hendrix *et al*., 2000). Newly incorporated morons can be recognized by the significant difference in GC content compared to surrounding genes. A moron that benefits the phage is fully integrated into its genome over time (Hendrix *et al*., 2000; Hendrix, 2002) .

Phage genomes are highly mosaic (Brüssow & Desiere, 2001). This means that each phage genome can be described as being composed of multiple, unique and exchangeable functional units from a common gene pool. These functional units are referred to as 'modules' and examples include head, tail and structural genes (Hendrix, 2002; Brüssow *et al*., 2004). Modules differ in size, rate of exchange and genomes in which they are found (Hatfull, 2008).

The recombination events that produces the high level of mosaicism observed in phages is driven by lateral gene transfer and involves homologous, and less often heterologous, exchange of genetic material between phages of different sizes, morphology and host range (Hendrix *et al*., 1999). Two models have been proposed to explain how this recombination occurs.

The first model proposed that gene junctions contain short conserved boundary sequences that serve as targets for the catalysis of homologous recombination by phage-encoded or host-encoded recombinases (Susskind & Botstein, 1978). Although a number of potential boundary sequences have been reported (Clark *et al*., 2001), their functions have not been satisfactorily

described and given that these potential recombination targets are not widespread, it is highly unlikely that the majority of recombination events follow this model.

The second model suggests that recombination events are random, frequently leading to the disruption of essential gene functions and mostly resulting in the production of non-viable phages which are rapidly eliminated (Hendrix, 2002). However, multiple recombination events may sometimes result in productive genetic exchange leading to the incorporation of functional modules of the right size into a genome. This viable combination of genes may result in the assembly of a 'new' phage. A study suggested that phage-encoded RecE or RecT- like recombination systems (Martinsohn *et al*., 2008) or CRISPR sequences (Sorek *et al*., 2008) may be involved in generating mosaicism through homologous recombination.

An alternative approach to understanding phage mosaicism is to compare predicted phage protein sequences. This approach is useful because many phages do not share any similarity at the nucleotide level, hence protein sequences are better predictors of evolutionary ancestry and mosaicism. Phylogenetic reconstructions using protein sequences have shown that different genes, gene clusters and or gene segments all have different ancestry and therefore represent individual modules within a mosaic genome (Pedulla *et al*., 2003; Hatfull *et al*., 2006; Liu *et al*., 2006).

Phage diversity is high enough to significantly limit the broad application of phylogenetic comparisons. Phamerator (Figure 1.3), a bioinformatic tool, was recently developed to include

genomes that do not belong to the same phylogenetic cluster in the analysis of phage mosaicism (Cresawn *et al*., 2011).

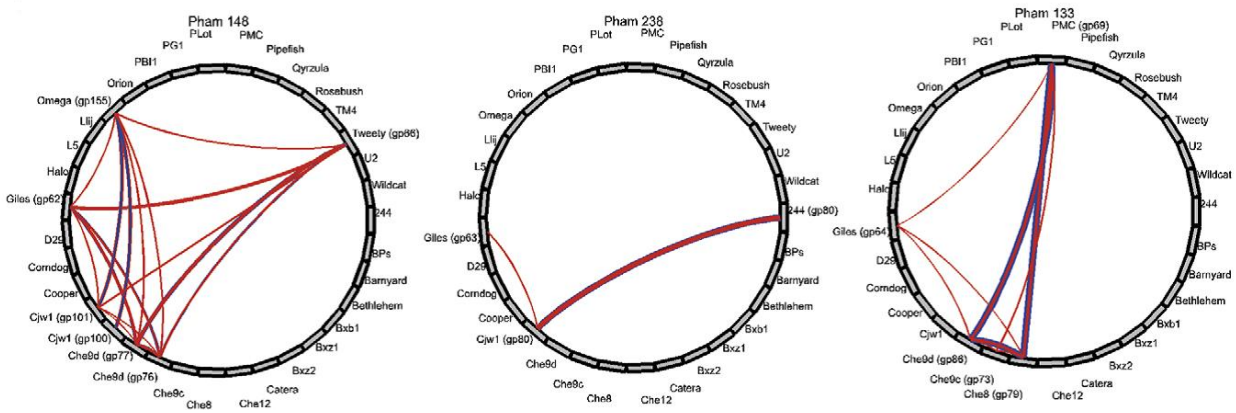

Figure 1.3: Phamily circles of three consecutive genes (62–64) of mycobacteriophage Giles. Each circle corresponds to three different Phams representing group of related proteins; and the thickness of the lines connecting each phamily constituent is related to the strengths of the relationships. Each of the three Giles genes clearly has a different evolutionary history (Hatfull, 2008).

Generally, mosaicism in phage genomes is so extensive that any attempt to define phylogenetic relationships of whole genomes would be limited to a few very closely related phages. This phenomenon forms the basis of identifying closely related phages using genes such as terminase and g20 and attempting to identify potential bacterial hosts.

1.6 PHAGE-HOST INTERACTIONS

Phages do not possess the ability to move independently hence, for infection, they depend on random contact with specific receptors on the bacterial host (Gabashvili *et al*., 1997; Moldovan *et al*., 2007). They bind to receptors such as proteins, teichoic acids, lipopolysaccharides or flagella on the host cell surface (Riede *et al*., 1985; Rakhuba *et al*., 2010). The specificity of this action means that most phages have a narrow host range (Carlton, 1999; Ohno *et al*., 2012).

Contact with the specific receptor is mediated *via* tail fibres or similar structures for non-tailed phages. For example, in a number of susceptible gram-negative bacteria, the phage uses the pilus to guide contact with the cell. When the pilus is retracted by the bacterium, bound virions are translocated to the outer membrane of the bacterium (Romantschuk & Bamford, 1985; van Duin, 1988; Webster, 1996), a process which is thought to enable penetration of the possible slime layer on the bacterial surface (Romantschuk & Bamford, 1985). Examples of such phages include the filamentous ssDNA viruses (e.g., M13, fd and fl) (Webster, 1996), icosahedral ssRNA phages of *E. coli* (e.g., Qβ and MS2) (van Duin, 1988) and the enveloped phage ϕ6 of *Pseudomonas syringae* (Romantschuk & Bamford, 1985). The filamentous phages interact specifically with the tip of the host cell F-pilus using the minor coat protein (pIII) (Jacobson, 1972) while the icosahedral ssRNA phages and phage ϕ6 attach to the side of the pili (Bamford *et al*., 1976; van Duin, 1988).

The phage attachment process can be reversible, referred to as desorption, where the phage simply falls off, or irreversible, upon which infection occurs (Schwartz, 1975). Upon irreversible attachment to a susceptible host, the bacteriophage generally pierces the cell membrane and injects its packaged nucleic acid into the cell in a specialized manner (Mayer, 2007; Chatterjee & Rothenberg, 2012). The nucleic acid is the only phage component that enters the cell, while the viral capsid remains attached to the bacterial cell surface (Poranen *et al*., 2002) and the effectiveness of the injection of genetic material is thought to be enhanced by the presence of ATP in the tail, if present (Fiers *et al*., 1976; Mayer, 2007).

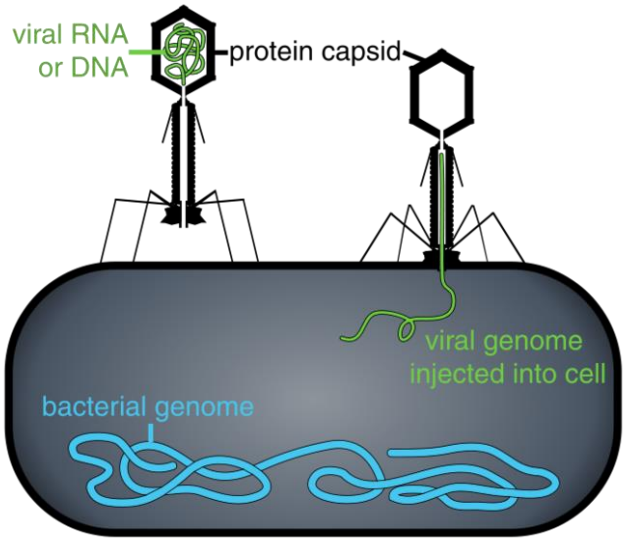

Figure 1.4: Diagram of how some bacteriophages infect bacterial cells. (https://www.boundless.com/microbiology/textbooks/boundless-microbiology-textbook/viruses-9/culturing-viruses-120/batch-culture-of-bacteriophages-619-5182/ [Accessed: 11-03-2016]).

Tailed dsDNA bacteriophages deliver their genome into the cytoplasm through a circular portal complex. The portal is located at one of the vertices of the capsid, connecting the tail to the capsid (Valpuesta & Carrascosa, 1994). The tail is involved in host recognition and attachment, which facilitates genome delivery into the cell (Poranen *et al*., 2002). The piercing action of the tail is thought to induce some permeability changes in the plasma membrane leading to the formation of a channel or pore within the host envelope (Dreiseikelmann, 1994). The pore for phage genome delivery across the envelope differs for phages. For example, λ phage induces the formation of a channel using cellular proteins (Roessner & Ihler, 1986), while phage T5 encodes a pore-forming protein (pb2), as a part of its tail, and forms a hydrophilic transmembrane channel within lipid membranes (Feucht *et al*., 1990). Phage T7, encodes proteins (gp14, gp15, and gp16) which are ejected from the virion through the tail and probably mechanically create a channel across the cell envelope (Molineux, 2001). It has been suggested

that phage T4 creates a pore by inducing the fusion of the outer and plasma membranes (Tarahovsky *et al*., 1991).

The mechanism of genome delivery has also been studied for non-tailed dsDNA phages. For example, phage PRD1 has no tail but possesses an envelope, which is a lipid bi-layer between its genome and capsid. On contact with a receptor, one of the capsid vertices releases a receptor-binding spike, causing the formation of a 14 nm opening within the receptor (Butcher *et al*., 1995; Rydman *et al*., 1999). Part of the phage envelope is transformed into a tubular structure that penetrates the capsid through the opening at the vertex. Viral membrane proteins assist this "tail-tube" structure in penetrating the cell envelope, creating a link between the interiors of the cell and the virus through which the genome is delivered into the cytoplasm (Poranen *et al*., 2002).

The fate of the phage capsid after successful genome delivery has been studied. The capsid may fall off, but for some phages, it has been shown to fuse with the bacterial outer membrane afterwards. For example, phage ϕ6 possesses an envelope and contains extension oligomeric units composed of two proteins, the receptor-binding spike, P3, and the fusogenic transmembrane protein, P6, which anchors the spike (Bamford *et al*., 1987; Overbaugh *et al*., 2001). P6 is activated once the spike protein P3 completes its interaction with the bacterial pilus (Romantschuk & Bamford, 1985; Bamford *et al*., 1987). This initiates the pH-independent fusion of the phage envelope with the bacterial outer membrane (Bamford *et al*., 1987).

In non-enveloped phages such as fd, a filamentous phage of *E. coli*, the major capsid protein (pVIII) is associated with the bacterial plasma membrane during both virus entry and assembly (Webster, 1996). The insertion of the coat protein during the entry process logistically resembles the insertion of the viral envelope into the host membrane. Phage fd also uses the

bacterial pilus for attachment but requires an obligatory co-receptor, the C-terminal domain of the periplasmic protein TolA, for infection (Riechmann & Holliger, 1997). The interaction with the co-receptor triggers the virion uncoating, and the insertion of the viral major capsid protein into the bacterial plasma membrane (Click & Webster, 1998). The fusion of the phage capsid is a carefully mediated process which helps to preserve the integrity of the cell membrane and avoid leakage of cellular components from the junction of the viral and host components (Poranen *et al*., 2002). Upon successful genome delivery, a phage can enter one of four life cycles which are briefly described below:

1.6.1 Lytic cycle

The *E. coli*-infecting T4 phage (Figure 1.5) is a classic example of a phage capable of undergoing the lytic but not lysogenic cycle. It is one of the largest known phages with a length of approximately 200nm and 90nm width and possesses hollow tail fibres adapted for injecting genetic material in a hypodermic syringe- like manner (Rabbani *et al*., 2004). T4 phage encodes about 200 genes which have been widely studied (Knipe & Howley, 2001).

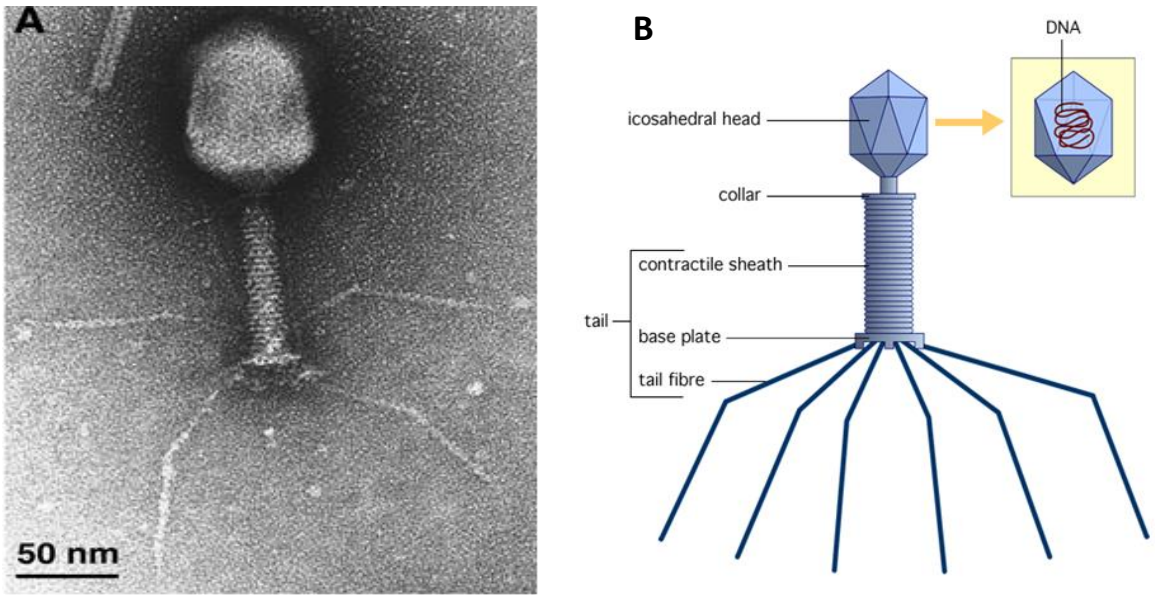

Figure 1.5: (A) An electron micrograph of T4 phage showing the morphology with distinct tail fibres (Miller *et al*., 2003) (B) A descriptive model of bacteriophage T4 (bugs.bio.usyd.edu.au [Accessed: 25-04-2013]).

The life cycle of T4 phage is summarized as follows: host recognition is achieved through contact between the tail and tryptophan or liposaccharide receptors on *E. coli*'s surface (Goldstein *et al*., 1984; Hu *et al*., 2015). Once contact is made, the tail transmits a signal to the capsid to facilitate both the opening of the channel through which DNA moves and the genome is injected. The baseplate harbours some short tail fibres (STFs) and six long tail fibres (LTFs) and the LTFs can be divided into proximal and distal half-fibres which are angled at about 20 degrees from each other (Cerritelli *et al*., 1996; Ward *et al*., 1970). The LTFs are able to exist in two conformations: retracted (up) or extended (down) state (Conley & Wood, 1975). It was initially thought that the fibres were extended when the phage requires rapid adsorption to a host, and retracted when unfavourable conditions such as low temperature, low pH and low ionic strength, occurred, however, research has shown that for infection, each LTF exists in a state of dynamic equilibrium between retraction and extension (Kellenberger *et al*., 1996; Hu *et al*., 2015).

The collar (Figure 1.5B) and whiskers (not shown) are formed by the assembly of twelve copies of the *wac* protein just below the head-tail junction of the phage (Dickson *et al*., 1970; Yanagida & Ahmad-Zadeh, 1970; Coombs & Eiserling, 1977; Fokine *et al*., 2013). The *Wac* proteins attach to the hinge connecting the proximal and distal half-fibres, causing the proximal half-fibre to attach to the baseplate protein gp9. This attachment involves the binding of six gp9 trimers to the upper edge of the baseplate (Terzaghi *et al*., 1979; Leiman *et al*., 2010). This coaxial attachment of the LTFs to the C-terminal domain of gp9 allows the LTFs to orientate correctly with respect to the capsid and contractile sheath of the phage (Leiman *et al*., 2004; Kostyuchenko *et al*., 2005).

The LTFs, which have become extended, bind to the *E. coli* B lipopolysaccharide (Prehm *et al*., 1976) or outer membrane protein C (OmpC for K-12) (Montag *et al*., 1990) causing the transition of the baseplate from a hexagonal to star conformation, and increasing the diameter from 40 to 60 nm (Simon & Anderson, 1967a; Simon & Anderson, 1967b). This conformational change also unpins the STFs from the baseplate, causing them to rotate downwards and bind to the A-KDO region of the lipopolysaccharide (LPS) (Riede *et al*., 1985; Leiman *et al*., 2010). This destabilizes the metastable conformation of the baseplate, causing it to contract, and forcing the tail tube into the cytoplasm of the cell for genome injection in a syringe-like manner (Hershey, 1952; Stent, 1963). Cell wall degradation is critical to the entry of the tail tube into the cytoplasm and this is achieved by the presence of trimers of gp5 as the needle and the protein has been shown to possess lysozyme activity (Mosig *et al*., 1989). Gp5 is thought to dissociate and diffuse away as the tail tube successfully enters the cytoplasm (Kao & McClain, 1980). In this cycle, the injected genetic material exists as a separate entity within the cytoplasm of the host (Tarahovsky *et al*., 1991; Rossmann *et al*., 2004).

The genome of phage T4, approximately 170 kb dsDNA, contains several origins of replication (*ori*s) which can facilitate multiple rounds of DNA synthesis, and encodes proteins such as replicative helicase, topoisomerase and primase, which are involved in the replication process (Kreuzer & Brister, 2010). Seven putative origins of replication have been identified (named *oriA* through *oriG*) and at least five, *oriA*, *oriC*, *oriE*, *oriF*, and *oriG*, have been shown to be involved in early replication (Brister & Nossal, 2007; Brister, 2008). Using these *ori*s, the phage is able to initiate the first of two stages of replication; autonomous bidirectional replication. Studies of *oriF* and *oriG* have shown that they both contain an AT-rich downstream unwinding element (DUE) (Menkens & Kreuzer, 1988; Carles-Kinch & Kreuzer, 1997) and a middle-mode promoter which contains a binding site for the viral transcription factor MotA (Hinton *et al*., 2010). When transcription is initiated from the *oriF* MotA-dependent promoter, persistent R-loops are formed in the DUE region and this leaves the non-template strand hypersensitive to ssDNA cleavage (Carles-Kinch & Kreuzer, 1997; Belanger & Kreuzer, 1998). These R-loops are thought to be cleaved by viral RNase H, generating free 3'-OH ends used for priming the synthesis of the leading DNA strand by the host RNA polymerase (Carles-Kinch & Kreuzer, 1997; Belanger & Kreuzer, 1998). The phage encoded gp61 primase primes replication of the discontinuous lagging strand (Liu & Alberts, 1980; Nossal, 1980; Hinton & Nossal, 1987). A few minutes post-infection, the host RNA polymerase recognition site is altered in preparation for the expression of late phage genes, an action which also turns off the MotA-dependent middle mode promoters of *oriF*, and *oriG* (Hinton *et al*., 2010). This forces the phage to switch to stage two, recombinant-dependent replication (RDR), where it transcribes late genes using its own machinery and uses recombinant intermediates for the leading strand DNA synthesis of the instead of RNA primers (Goldstein *et al*, 1984).

Genome attributes of the T4 phage, namely its terminal redundancy (two terminal regions of each genome having the same sequence) and circular permutation (a consequence of random

cleavage of the circularized genome), make RDR possible (Streisinger *et al*., 1967). Terminal redundancy results in each 3′ ss end generated from bidirectional replication being homologous to the 5' end of the DNA strand eventually leading to the generation of long DNA concatemers which is required for DNA packaging into new phage heads that will be assembled later (Kreuzer, 2000).

When the expression of late genes reaches a threshold, viral assembly is activated by helper proteins starting with the viral base plate which is attached to the tail and tail fibre proteins (Wommack & Colwell 2000; Snyder & Champness 2007). A mature phage capsid is also assembled and DNA is packaged using a simple mechanism- until the head is full. The DNA concatemer generated from recombinant-dependent replication is resolved by endonuclease VII (Mizuuchi *et al*., 1982) and packaging is initiated randomly along the phage genome (Kalinski & Black, 1986). Headful packaging ensures that each phage head contains approximately 103% of the genome length, making its terminally redundant by default (Streisinger *et al*., 1967).

The release of mature phage particles is facilitated by phage encoded enzymes. The inner membrane of the bacterial host is punctured by the enzyme holin, an action which allows endolysins and lysozyme access to the peptidoglycan cell wall causing its degradation. This significantly weakens the membrane thus allowing either the outflow of cellular content or influx of fluid from the external environment into the cell, thus bursting it and liberating viral particles (Cann, 1994). This results in cell lysis and release of newly assembled viral particles into the environment (Goldstein *et al*, 1984).

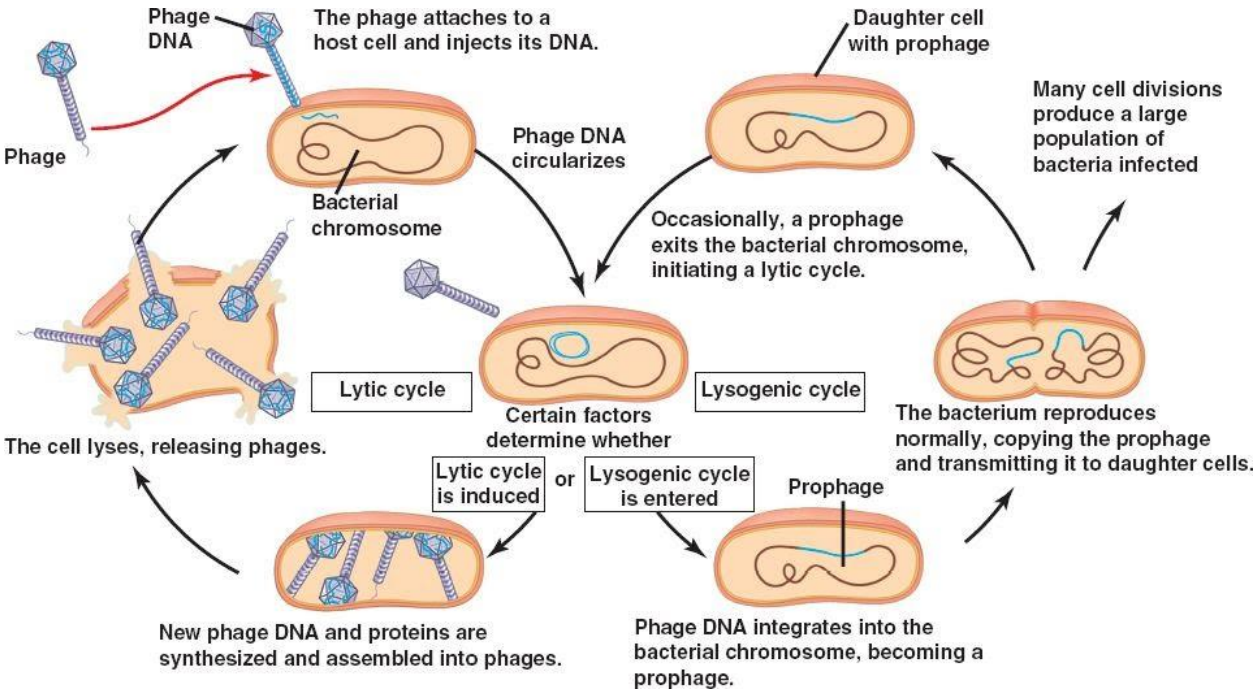

Figure 1.6: Schematic of lytic and lysogenic life cycles of phages. Lytic phages immediately enter a productive cycle, hijacking the bacterial cell machineries to replicate the phage genome, synthesize capsid and tail proteins, packaged the phage genome into progeny phage particles and release new virions via bacterial lysis. Temperate phages enter a lysogenic cycle, in which the phage genome is integrated into the bacterial genome and persists in a latent state that does neither promotes cell death nor the production of phage particles. Prophages are replicated together with the bacterial host genome during cell division but switch into lytic cycle upon exposure to DNA damage. (https://sites.google.com/site/cundiffbiologyportfolio/home/science-as-a-process/viral-life-cycles) [Accessed: 4-1-2016].

1.6.2 Lysogenic cycle

In the lysogenic (temperate) cycle, the phage genome is incorporated into the genome of the host and is referred to as prophage. As the host replicates, it passes along the viral genome to its daughter cells (Canchaya *et al*., 2004). The viral genome remains dormant until an induction event such as DNA damage, lack or excess of nutrients, or change in growth condition, causes conversion to the lytic cycle where more phage particles are assembled and released (Oppenheim *et al*., 2005).

A feature referred to as lysogenic conversion has been observed in a number of lysogenic phages. Naturally, the phage genome is incorporated into the host's genome but not expressed until induced. However, in some instances, a region of the phage sequence may be expressed alongside the host genes leading to phenotypic changes in the host (Canchaya *et al*., 2004). Such changes may include making the host cell membrane more or less permeable to other phages which can alter its pathogenicity (Keen, 2012). This phenomenon has been observed in *Corynebacterium diphtheriae* where the toxin identified as the main virulence factor is encoded in the genome of phage β (Mokrousov, 2009). Other examples include *Streptococci* where the pyogenic exotoxin responsible for the skin rash observed in scarlet fever is only expressed when the bacterium is a lysogen (Todar, 2001). Staphylococcal enterotoxin, panton-valentine leucocidin (PVL), exfoliative toxin (ETA) and staphylokinase (SAK) are also phage encoded in Staphylococci (Riley, 2006).

A widely studied example of a lysogenic phage is Lambda (Figure 1.7). The Lambda phage is a tailed phage capable of undergoing both lytic and lysogenic cycles and has been the basis of several studies that have increased our understanding of gene functions and regulation, DNA replication and virion assembly (Wommack & Colwell, 2000; Rajagopala *et al*., 2011).

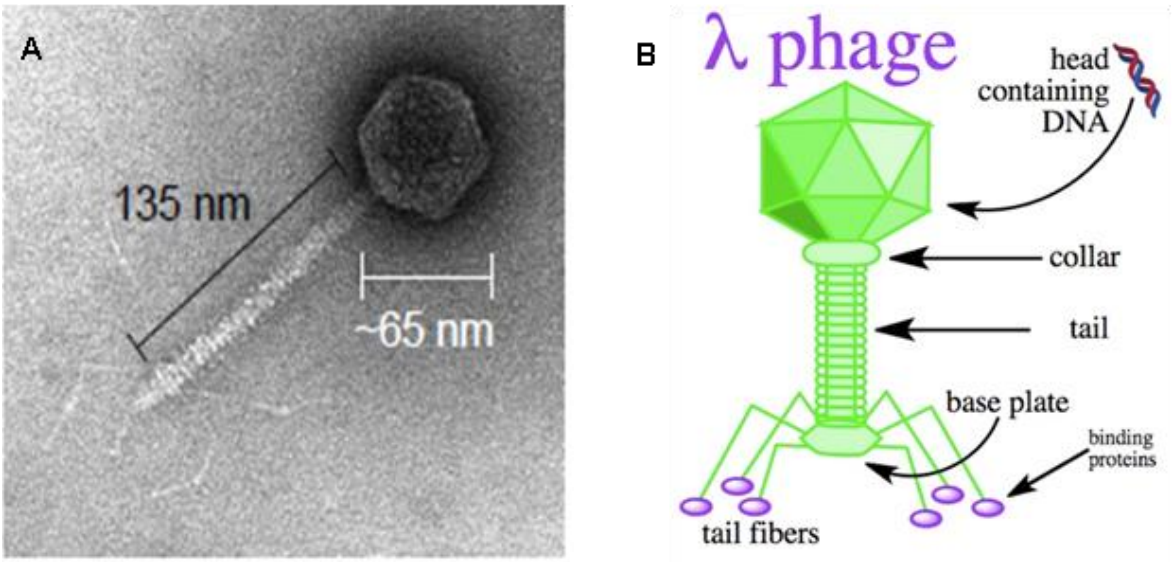

Figure 1.7: (A) An electron micrograph of λ phage showing the morphology with distinct tail fibres (Rajagopala *et al*., 2011) (B) A descriptive model of λ bacteriophage (en.wikibooks.org [Accessed: 25-04-2013]).

Detailed studies on this phage have focused on characterizing the genetic switch that determines whether the phage is incorporated into the host genome, thus becoming a prophage, or new viruses are assembled leading to cell lysis (Court *et al*., 2007).

In *E. coli*, lambda phage attaches and injects its DNA through the interaction of *J* protein found at the tip of its tail and the *lamB* gene product from the host. Host RNA polymerase binds to viral promoters *pR* and *pL* leading to bi-directional transcription (Werts *et al*., 1994).

Viral genome integration is driven by *int* (integrase) protein which simultaneously binds to both the *attB* site on the bacterial genome and the *attP* site on the viral genome. Site- specific integration implies that a host genome cannot have more than one copy of lambda gene sequence under normal circumstances (Goldstein *et al*., 1984; Rajagopala *et al*., 2011). When induced, the lytic cycle of lambda follows a similar pattern to T4 phage (Figure 1.8).

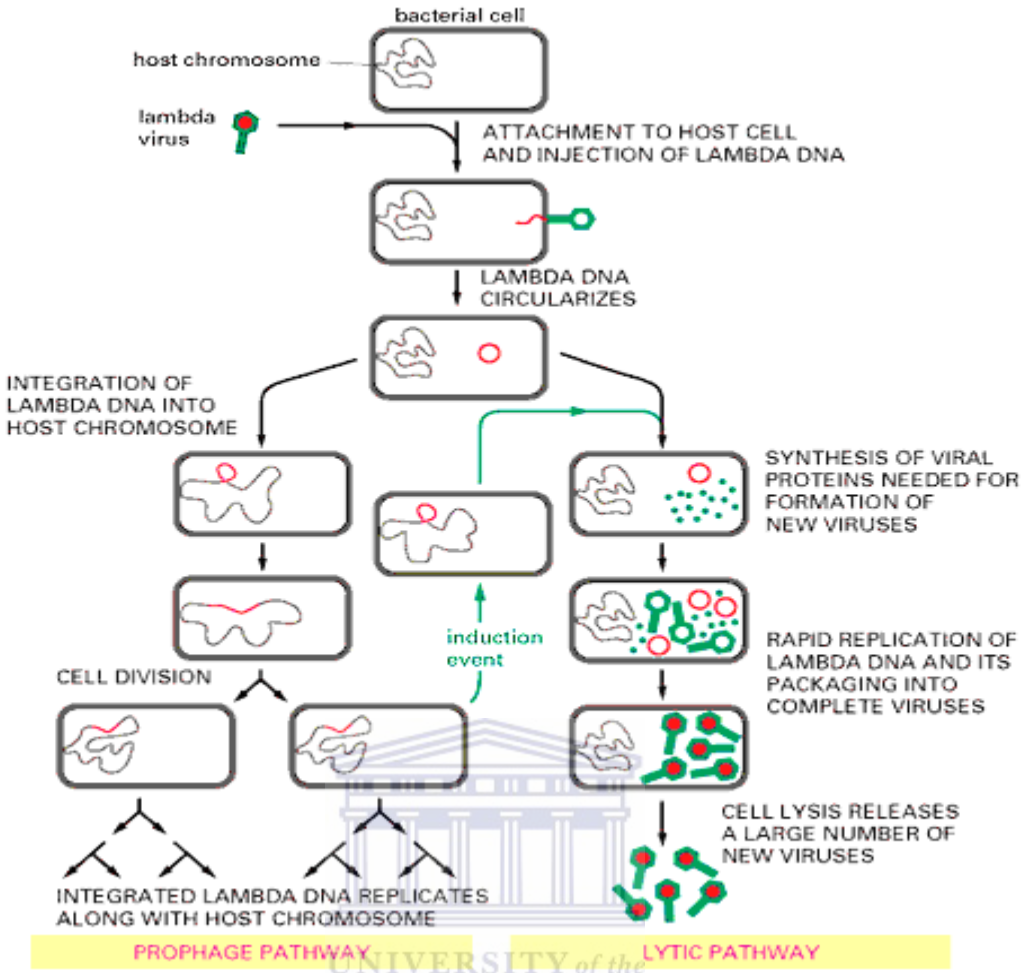

Figure 1.8: The lysogenic life cycle of lambda phage with a description of the effect of an induction event (Todar, 2001).

1.6.3 Pseudolysogeny

This simply refers to phages that are neither lytic nor lysogenic under prevailing conditions (Łoś & Węgrzyn, 2012). It is an unstable condition where the phage fails to actively replicate and produce new viral particles nor establish itself as a prophage (Figure 1.9). Pseudolysogeny predominantly occurs under nutrient deprivation where the host bacterial cell is incapable of supporting either DNA replication or protein synthesis (Feiner *et al*., 2015). The phage genome remains in a non-integrating and non-replicating pre-prophage state, where it is referred to as an episomal phage, until a favourable nutritional state is restored, at which point it can either

enter a lytic or lysogenic life cycle (Miller & Day, 2008). The pseudolysogenic preprophage does not replicate, therefore, if the bacteria undergoes cell division while the nutrient deprivation state persists, the phage is inherited by only one daughter cell (Feiner *et al*., 2015).

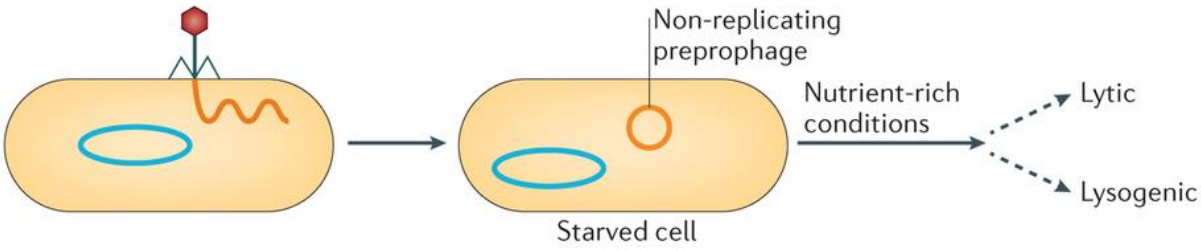

Figure 1.9: Schematic representation of Pseudolysogeny (Feiner *et al*., 2015).

Classifying phages as pseudolysogenic can be difficult, as the phenomenon relies on the physiological state of the host cell. Bacterial cells are known to have variable growth rates under optimized laboratory conditions compared to natural environments. The latent time for a lytic phage might therefore be extended due to lack of some nutrients, leading to slower assembly of new phage particles. Such a phage might be wrongly classified as pseudolysogenic (Łoś & Węgrzyn, 2012). Also, while the regulation of lysogeny depends on the stability of the repressor gene and its sensitivity and reaction to different inducing agents (Wegrzyn & Wegrzyn, 2005), the process is more complicated in episomal prophages (Yarmolinsky, 2004). The typically lytic phage T4 is a classic example of pseudolysogeny. It is capable of producing pseudolysogens when infecting a starved, non-growing bacterial host (Kutter *et al*., 1994). It was also demonstrated that activity of the *rI* gene was responsible for pseudolysogeny in phage T4 when infecting *E. coli* cells cultured under slow growing conditions in a chemostat at 25°C (Los *et al*., 2003). Other examples of phages capable of pseudolysogeny include phages F116 and UT1 of *Pseudomonas aeruginosa* (Ripp & Miller, 1998).

1.6.4 Phage resistance

This is similar to pseudolysogeny in that the phage successfully attaches to the host and injects its genetic material but does not replicate. However, while pseudolysogeny occurs as a result of nutrient deprivation as described above, phage resistance is a bacterial defence mechanism. At least two strategies have been identified: CRISPR system (acquired immunity) and abortive infection (innate immunity).

The CRISPR (Clustered regularly interspaced short palindromic repeats) system was first discovered in *E. coli* (Ishino *et al*., 1987) and according to the CRISPR database (CRISPRdb), the system exists in about 45% of bacteria and 85% of archaea (Grissa *et al*., 2007). It has however been suggested sampling bias might be responsible for the difference in prevalence (Rath *et al*., 2015). The CRISPR system involves a mechanism where over time, the bacteria recognizes specific sequences from phages it is susceptible to, and synthesizes short nucleases to degrade the phage genetic material once it is within the cell (Szczepankowska, 2012; Rath *et al*., 2015). The system comprises of a range of short repeat sequences (usually 21 bp to 48 bp) separated by spacers (26 bp and 72 bp) (Jansen *et al*., 2002; Bolotin *et al*., 2005; Grissa *et al*., 2007). While CRISPR sequences are usually found on chromosomal DNA, spacers are fragments of viral and plasmid nucleic acid (Rath *et al*., 2015) and this forms the basis of phage resistance (Bolotin *et al*., 2005; Mojica *et al*., 2005; Pourcel *et al*., 2005). Bacteria and archaea acquire new spacers over time and are able to develop resistance to new phages. CRISPRs can therefore be considered as an ancestral record of the viruses the cell has encountered. The process of destroying foreign genetic material can be divided into three stages: i. Adaptation; where newly acquired spacers are inserted into the CRISPR locus (Barrangou *et al*., 2007; Wei *et al*., 2015). ii. Expression; where *cas* genes are expressed, CRISPR sequences are transcribed into precursor CRISPR RNA which are converted into mature CRISPR RNA by Cas proteins, and accessory factors such as RNase III, tracriRNA and some currently unknown factors

(Garneau *et al*., 2010; Deltcheva *et al*., 2011; Jinek *et al*., 2012; Wei *et al*., 2015). iii. Interference; where mature CRISPR RNA and Cas proteins combine to destroy the target nucleic acid (Bikard & Marraffini, 2013; Goldberg *et al*., 2014; Rath *et al*., 2015). This ensures the phage is not able to replicate and the bacterial cell remains intact.

Abortive infection involves programmed cell death, where the bacterium self-destructs rather than propagate new phages (Dy *et al*., 2014). While the molecular basis is not fully understood, the process is thought to be a method of containing phage infection within a bacterial population, with the infected bacterium essentially committing suicide, ensuring that the phage genetic material is lost and new viruses are not assembled (Chopin *et al*., 2005; Dy *et al*., 2014).

1.7 POTENTIAL BENEFITS TO STUDYING PHAGE-HOST INTERACTIONS

1.7.1 Better understanding of bacterial turnover in the environment and biogeochemical cycles

The bacteriolytic action of phages has been shown to play an important role in the transfer of matter and energy within ecosystems (Paul *et al*., 2002). The result of the lytic cycle includes new viral particles and cellular debris (Paul & Sullivan, 2005). The fate of this cellular debris has been studied in relation to ecology and biogeochemical cycling. Cellular debris is made up of polysaccharides, proteins, colloids and cell fragments among other products. These products contribute to what is referred to as dissolved organic matter (Middelboe, 2008).

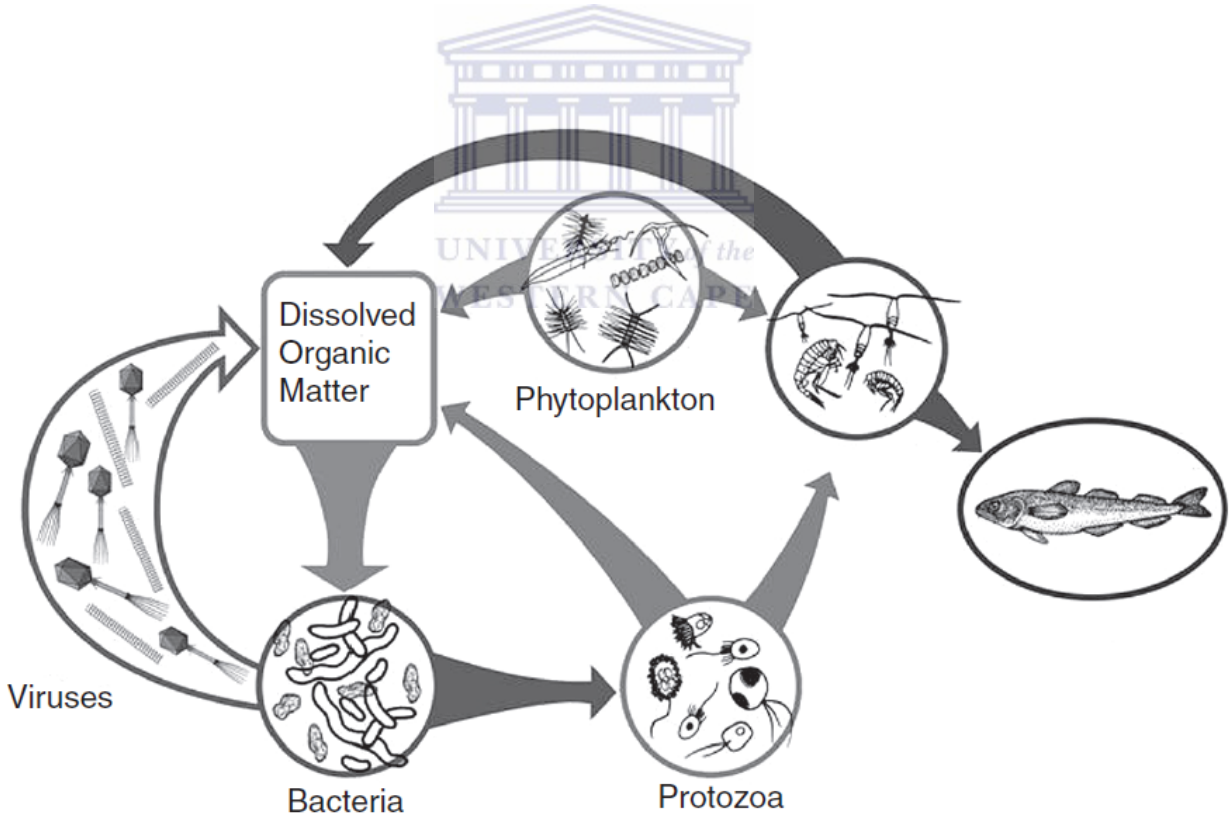

Figure 1.10: A model of the proposed concept of biogeochemical cycling in a marine food web. Organic matter lost from each trophic level in the form of dissolved organic matter (DOM) can be utilized by bacteria. Viral infection and cell lysis play a significant part in recycling consumed dissolved organic matter (Middelboe, 2008).

The physical and chemical properties of dissolved organic matter released into an environment can be used to explain some of the properties of that environment. For example, cellular components may contain polymers which contribute to the viscosity of sea water and affect the ability of microscopic organisms to adhere to the water surface via surface tension (Proctor & Fuhrman 1990).

Viral actions have also been suggested to have a significant effect on global climate by enhancing cloud nucleation. The release of dimethyl sulphide (DMS), as a by-product of the infection and lysis of marine phytoplankton by viruses, has been shown to contribute to cloud nucleation which involves the formation of particles on which water can be converted from vapour to liquid in the atmosphere. Other major contributors to this event include phages known to infect species such as *Micromonas pusilla* and *Emiliania huxleyi* (Hill *et al*., 1998; Brussaard *et al*., 2000).

1.7.2 Finding novel enzymes through phage metavirome sequencing

Phage encoded enzymes are routinely used in molecular biology. Examples include T4 DNA polymerase and T4 ligase which are involved in viral replication. The exit process is also mediated by enzymes such as endolysins and holins. These two classes of proteins are of significant interest in scientific studies, especially in industrial processes that involve the extraction of cellular products and the development of new antibiotics (Paul *et al*., 2011).

The ubiquitous nature of phages and bacterial hosts attests to the adaptability of these biological entities. Such adaptability may involve modifying and optimizing the phage replication process using enzymes with multiple activities. For example, a thermostable DNA polymerase which also has reverse transcriptase activity was recently discovered while screening a metaviromic library constructed from Octopus hot spring (93^0C) (Moser *et al*., 2012). Finding similar

enzymes will greatly improve currently available molecular biology tools and the optimization of industrial processes.

1.7.3 Development of phages as a genetic tools

Phages have been studied, modified and used as genetic tools largely due to their inherent ability to transport genetic material. Applications of phages in genetic engineering include:

1.7.3.1 Phage display

This is a very useful molecular technique for screening for novel polypeptides. The concept of phage display involves fusing the gene that encodes the polypeptide of interest with genes for the protein coat of the phage. This results in the display of the desired polypeptide on the surface of the phage particle, a process referred to as tagging (Smith, 1985; Sidhu, 2000). From phages displaying the desired polypeptide, libraries can be constructed and screened for peptides with the required degree of specificity and affinity for target proteins. This technique is particularly useful in the pharmaceutical industry for understanding the pattern of molecular recognition of compounds under consideration and antibody screening (Sidhu, 2000). Examples of phages currently used in phage display include lambda phage, filamentous phage M13 of *E. coli*, and T7 phage (Benhar, 2001). Recent advancements have led to the development of products such as NovoPhage™ Phage Display monoclonal antibody capable of creating antibody libraries each with an abundance of over10^{10} plaque forming units and a gene coverage greater than 85% (www.novophage.com [Accessed: 25-11-2016]).

1.7.3.2 Phage typing

Typing of bacterial strains can be carried out by exploiting the specificity of phages. This involves using the sensitivity patterns of bacterial cells to specific phages as a basis of identification (Watson & Eveland, 1965; Clark & March, 2006). For the identification of an unknown bacterial strain, its lawn is provided with different phages and appearance of plaques is monitored. If plaques appear, lysis of the bacterial population is implied and the specific strain can be determined (Clark & March, 2006).

Fluorescent proteins which are only expressed after infecting a bacterial host can also be covalently attached to phage protein coats to identify a particular strain. Alternatively, a nucleic acid intercalating fluorescent dye, detectable after phage infection, can be used (Hennes *et al*., 1995; Goodridge *et al*., 1999a).

1.7.3.3 Targeted gene delivery

Gene delivery is the process of introducing foreign genetic material into a cell to modify its metabolism and function. Successful delivery of genetic material is key to phage replication and has been exploited in biotechnology. The ability to engineer phages to accept foreign DNA and express foreign proteins can be harnessed to ensure it delivers a gene sequence of interest to specific cell types (Clark & March, 2006).

A phage to be used as mammalian vector is imparted with cell specificity by linking it with a ligand that is known to specifically target a receptor on the cell of interest (Larocca *et al*., 2002). Ligands could be natural such as growth factors or an antibody against the surface receptor (Larocca *et al*., 2002).

Using a phage is potentially better than animal viral vectors because of the ease of producing high titres in bacterial cultures and its high stability under harsh conditions such as unfavourable pH and the presence of proteolytic enzymes (Larocca *et al*., 2002).

1.7.3.4 Phagemids

Phagemids are plasmids containing multiple origins of replication (*ori*): typically, ColE1 (for double stranded replication) and another from the filamentous phage f1. In phage f1, the *ori* functions as a signal for the initiation and termination of the synthesis of new viral strands (Dotto *et al*., 1984). Phage f1 is single stranded, therefore, its *ori* makes phagemids capable of single stranded DNA replication (Dotto *et al*., 1984; Endemann & Model, 1995). The f1 *ori* also allows for packaging into phage heads instead of transformation (Endemann & Model, 1995). Examples of phagemids include pUC18 and pUC19 and pBS vectors (Vieira & Messing, 1987; Short *et al*., 1988).

1.7.3.5 Cosmids

Cosmids are hybrid plasmids containing one or two cohesive end (*cos*) sequences from phage lambda (Hohn & Murray, 1977). These sequences are approximately 200 base pairs long. Both strands of the cosmid are nicked by the phage terminase at *cosN* sites located 12 base pairs apart within the *cos* site. This results in a linearized cosmid with 12 base pair overhangs at both ends and is required for efficient DNA packaging (Feiss & Catalano, 2005). Between 25-35 kb of DNA can be packaged into cosmids. In the presence of an *ori* such as ColE1 or f1, cosmids are capable of replicating like plasmids, or otherwise packaged into phage capsids which enables the transfer of foreign DNA from cell to cell *via* transduction (Hohn & Murray, 1977; Collins & Hohn, 1978).

There are currently only two phage packaging kits (MaxPlax™ Lambda Packaging Extracts and CopyRight® v2.0 from Lucigen) which are both based on phage lambda. Lambda phage is specific to *E. coli* which has been shown to be limited in its ability to express a wide range of foreign genes (Simon & Daniel, 2011). Hence, screening libraries constructed this way result in the expression of about 40% of all gene products (Simon & Daniel, 2011).

1.7.3.6 Promoters

Phage promoters have been used in recombination and expression studies. The majority of these studies involve promoters from phages that infect *E. coli* such as T3 (Brown *et al*., 1986; Klement *et al*., 1990), T7 (Klement *et al*., 1990; Tabor, 2001), P_L (Cheng & Patterson, 1992) and P_R (Łyzeń *et al*., 2009). Other examples include the SP6 promoter from phage SP6 infecting *Salmonella enterica* serovar Typhimurium LT2 (Brown *et al*., 1986; Dobbins *et al*., 2004).

1.7.4 Phage therapy

This refers to the therapeutic use of a phage or a cocktail of phages to treat pathogenic bacterial infections (Sulakvelidze, 2011; Reardon, 2014). Therapeutic applications of phages have been investigated and a wide range of potential uses have been found (Table 1.2). However, these applications have unique challenges which must be overcome before widespread application can be feasible as summarized below:

Table 1.2: Potential applications of phages (Clark & March, 2006).

APPLICATION	ADVANTAGES	DISADVANTAGES
Decontamination of surfaces and/or environment	Can kill antibiotic-resistant bacteria	Not as broad spectrum as other decontaminants. Effective life span is limited.
Carcass treatment	Specific food-poisoning bacteria can be targeted	Potential regulatory problems if used with food products.
Plants	Few regulatory problems expected. Phage shouldn't be specifically inactivated Can treat specific diseases	Potential problems with non-specific phage inactivation due to environment
Aquaculture	Can potentially be delivered in feed Phage neutralizing antibodies do not seem to be produced in fish	Might need to administer phage continuously or relatively soon after infection More studies are needed.

Phage therapy in animals	Fewer regulatory hurdles Can reduce the use of antibiotics Multiple delivery routes have been tested	Needs to be proven in large-scale studies Probably not applicable for long-term treatments (e.g. where broad-spectrum antibiotics are used to enhance growth.
Humans: Oral	Easy to deliver Potential to deliver in food and/or drink. Highly specific, therefore will not damage gut flora	Might need to neutralize stomach acid High levels of endotoxin can be produced from lysed bacteria.
Humans: Intravenous	Spreads rapidly throughout the body Multiplies in vivo	Phage swiftly cleared by the immune system. Phage needs to be highly purified. Little clinical data available.
Humans: Topical	Can treat antibiotic-resistant infections. In vivo amplification of phage can enable them to penetrate into wounds. Phage can produce enzymes to break down biofilms enabling access to pathogens. Can be delivered conveniently (e.g. in creams, rinses and patches).	Might need continuous treatment with low levels of phage. Interference by the host immune system.

1.8 LIMITATIONS OF CURRENT SAMPLING METHODS

The standard procedure for isolating bacteriophages from environmental samples involves re-suspension (for solid samples) and filtration followed by mixing the phage fraction with bacterial cultures and plating out on solid media. Alternatively, low density species can be enriched for using liquid culture before assaying (Hyman & Abedon, 2009).

The success of this approach, however, depends on having an understanding of the metabolic requirement of the bacterial host. It is generally accepted that the majority of bacteria cannot be cultured under existing laboratory conditions (Hugenholtz, 2002). It's been estimated that less than one percent of marine bacteriophages can be detected using traditional isolation and propagation techniques (Chen & Lu, 2002). A consequence of this limitation is the bias observed in currently characterized phages, a significant number of which have relatively well studied bacterial hosts (Deschavanne *et al*., 2010).

The traditional detection technique also operates under the assumption that phages are usually smaller than 0.2 µm in diameter, and most studies use filters with these dimensions to exclude bacterial or eukaryotic cells. This size exclusion removes the possibility of discovering larger viral particles and as a result, less than two percent of sequenced phages have genomes greater than 200 kb. Statistical analysis suggests strongly the possibility of large-genome phages being under sampled (Claverie *et al*., 2006). To circumvent these challenges, researchers are adopting culture independent techniques to studying phages in natural environments.

1.8.1 The conserved gene approach

Previous studies have attempted to understand phage diversity by using conserved phage genes as markers. A number of these studies used Gene 20 as a diversity maker. Gene 20 (g20) is a homologue of the T4 phage portal protein and is conserved among T4- like myoviruses with hosts ranging from Proteobacteria to Cyanobacteria (Mann *et al*., 2005; Sullivan *et al*., 2008). It was initially identified as a potential marker because its evolution was likely to be restrained due to the geometric precision with which its protein product initiates capsid assembly (Hsiao & Black, 1978; Rao & Black, 2005). However, studies have demonstrated, using DNA fingerprinting and PCR with non-degenerate primers, the large extent of genetic variation across different environments over time which remains unrepresented in cultures (Frederickson *et al*., 2003). One of these studies sequenced g20 genes from 38 marine myophages isolated using *Synechococcus* and *Prochlorococcus* hosts. The majority of those sequences clustered into distinct groups representative of myophages specific to the host from which they were isolated from, but about 10% formed a separate cluster indicative of other potential primary hosts (Sullivan *et al*., 2008). While the study established some correlation between the host of isolation, g20 sequence and the host range, it also highlighted the relative frequency of exceptions. This suggested that phage portal proteins might not be good indicators of phage hosts or the optimal habitat of the phage itself (Sullivan *et al*., 2008).

Other single-locus studies have targeted genes like DNA polymerase, holins and terminase. While DNA polymerase genes have been found to be ubiquitous in the environment (Breitbart *et al*., 2004), it has been shown that their diversity even within a small geographical location is greater than previously thought (Labonté *et al*., 2009). Holins and terminases have also been successfully used as intra-genus markers (Liu *et al*., 2011; Shan *et al*., 2012). These studies have shown that closely related hosts and their phages are widely distributed and can be detected. However, this approach does not completely avoid culturing while studying phage

diversity and is still very dependent on PCR primers selected based on sequences from cultured hosts. There is currently no universal signature gene that can be used as phylogenetic markers for studying phage diversity (Edwards & Rohwer, 2005).

1.8.2 The metagenomic approach

Metagenomics aims to bypass the above-mentioned limitations and provide both qualitative and quantitative information about phages and bacteria present in an environment. Viral metagenomics involves the direct extraction and purification of viral particles from environmental samples, and using a range of molecular techniques and bioinformatic tools to determine the genomes, establish relative abundance and diversity, understand microbial interactions and identify potential hosts in a high throughput manner. Bioinformatic tools such as PHACCS (Angly *et al*., 2005), PHAST (Zhou *et al*., 2011) and Metavir2 (Roux *et al*., 2014) have been developed to analyse the large amount of metaviromic sequence data generated using this approach.

The metagenomic approach has a number of biases. Most studies have focused on analysing total metaviromic DNA from free living phages due to the relative ease of extracting metaviromic DNA compared to RNA (Mizuno *et al*., 2013; Hurwitz *et al*., 2013). They have also been limited to phages that remain stable during purification steps such as tangential flow filtration (TFF), polyethylene glycol (PEG) aggregation or caesium chloride gradient centrifugation (Kropinski, 2009; Hurwitz *et al*., 2013). Phages have different decay rates, therefore, storage under standard laboratory conditions, for example, 4°C, could potentially result in the loss of some novel species (Angly *et al*., 2006). Metaviromic DNA is also routinely amplified using whole genome amplification techniques which have been shown to preferentially amplify ssDNA and circular dsDNA (Hutchison *et al*., 2005; Yokouchi *et al*.,

2006). Another limitation is the size of currently available databases. No similarity to sequences on the database is found for a significant portion of metaviromic data (Hatfull, 2008).

Despite these drawbacks, valuable understanding of phage diversity has been derived using this approach. Studies targeting RNA viruses have supported the hypothesis that DNA viruses dominate the marine environment (A. Culley *et al*., 2006; Yassin & Mankin, 2007). Other studies have shown that local communities typically enrich for specific phage types (Desnues *et al*., 2011; Rosario *et al*., 2009). This has strengthened the suggestion that the type of nucleic acid (ssDNA vs dsDNA, DNA vs RNA) and composition of the viral community is a function of regional selective pressure and geographical location (Angly *et al*., 2006).

Also, metagenomics has enabled the expansion of phage diversity studies to a global scale. For example, the analysis of samples collected across four oceanic regions showed the presence of over 129,000 phage types in the British Columbian coastal waters (Angly *et al*., 2006). This study emphasized that viruses are capable of migrating across biomes leading to regional diversity being as high as global diversity (Sano *et al*., 2004). For example, phage P-SSM4, infecting *Prochlorococcus* was found at all sampling sites during the Sorcerer II Global Ocean Sampling Expedition (Williamson *et al*., 2008). This suggests that P-SSM4 plays a significant role in the distribution and abundance of one of the major components of picophytoplankton in oligotrophic oceans (Williamson *et al*., 2008).

Recent studies have been able to look at phage diversity, ecological impact and biotechnological products in extreme environments using metagenomic tools. Such extreme environments include in Hot Springs (Schoenfeld *et al*., 2008), hypersaline systems (Emerson *et al*., 2012), hot acid lakes (Diemer & Stedman, 2012), (hado)pelagic sediments (Yoshida *et*

al., 2013) and Namib Desert hypoliths (Adriaenssens *et al*., 2015). Such harsh environments are grossly understudied and are potential sources of novel phages and phage products.

1.9 DEVELOPING NEW PHAGE- HOST PAIRS ISOLATION TECHNIQUES

Technological advances and the improvement of currently available tools have led to the exploration of novel methods for identifying and isolating novel phage-host pairs. Two such techniques are explored in this study.

1.9.1 Reverse metaviromics

This involves working backwards from the analysis of environmental phage sequence data to the isolation of a phage and its host. Due to the high throughput nature of data generated from metaviromic sequencing, novel phages can be rapidly identified and evaluated *in silico* for potential hosts. This study sought to demonstrate the viability of this approach for the very first time (Chapter two).

1.9.2 Fluorescent labelling

A number of studies have used fluorescent tags as a means of identifying the bacterial host of phages (Egley & Breitbart, 2003; Ohno *et al*., 2012). Some of the experiments have involved tagging the phage particles with fluorochromes and attempting to track the infected host. This approach is described as viral tagging (Deng *et al*., 2012). Another approach that has been used is called fluorescence *in situ* hybridization (FISH) and involves detecting phage genes within the genomes of the bacterial hosts (Moter & Göbel, 2000; Strandberg & Enfors, 1991). This relies on the phage being lysogenic. Depending on the aim of the study, the fluorescence can

be detected using fluorescence microscopy or fluorescence activated cell sorting (FACS). These techniques are further expounded in Chapter four.

1.10 AIMS AND OBJECTIVES

The initial aim of this project was to develop a high throughput assay that combines fluorescence with flow cytometry for the identification and isolation of novel bacteriophages and the bacterial hosts simultaneously. The project also aimed to evaluate the use of next generation sequence data as a guide for the isolation of novel bacteriophage species from extreme environments.

The objectives of this study therefore included:

1. To develop a novel methodology for isolating phage-host pairs from the environment.
2. To explore the incorporation of metaviromic sequence data as a tool for the isolation of a novel phage.
3. To characterize any novel phage isolated.

2

USING METAVIROMIC SEQUENCE DATA TO GUIDE THE ISOLATION OF A NOVEL PHAGE

2.1 INTRODUCTION

Improvements in metaviromic tools have led to more studies into how the viral communities found in environments contribute to the ecology of those environments. Such tools include the development of advanced, less expensive and more accessible sequencing technologies, the expansion of currently available databases via the deposition and characterization of more genomes, the improvement of bioinformatic tools to handle complex data sets and better predict the metabolic requirements of microorganisms, and the increased scientific interest in studying microbial populations in nature rather than under laboratory conditions.

In this study, the aim was to use metaviromic sequence data to guide a bacterial isolation strategy for the eventual isolation of phages that infect them. This, in conjunction with whole genome sequence data, should provide a more informative and demonstrative overview of the nature of phage-host interactions. To the best of our knowledge, the concept of "from sequence to plate" has been demonstrated only twice and involved the isolations of a novel nitrogen fixing *Leptospirillum* bacteria from an acidophilic microbial community (Tyson *et al*., 2005) and an ectosymbiotic Nanoarchaeota ('*Nanopusillus acidilobi*') and its host (Acidilobus, a crenarchaeote) from a terrestrial geothermal environment (Wurch *et al*., 2016). The aim of this chapter was to apply a similar strategy for the isolation of phage-host pairs and refer to the process as "reverse metaviromics".

Due to the high throughput nature of data generated from metaviromic sequencing, potentially novel phages can potentially be rapidly identified and evaluated *in silico* for potential bacterial hosts. Given that phages usually harbour fragments of the genome of previous hosts within their genomes, potential matches can be identified using BLAST analysis and alignment with known bacterial genomes. Potential hosts are then selectively cultured using a combination of the metabolic requirements derived from genome information, and the physico-chemical properties of the sampling site. For example, a bacterial genome harbouring genes capable of

metabolizing xylan as a carbon source can be targeted by making it a limiting component. Once bacterial isolates are obtained in culture, they can then be challenged with the phage fraction from the environment which harbours the specific phage identified and the novel phage may be isolated.

An obvious limitation of this approach to phage isolation is its dependence on the availability of whole genome sequence of potential bacterial hosts on current databases from which metabolic requirements can be mined. Another limitation is that the majority of bacteria cannot be cultured which immediately restricts the number of novel phage species that can be isolated. However, this "reverse metaviromics" method provides some improvements on phage isolation by offering a more targeted approach to rapidly identifying and isolating novel phages compared to the classic method which involves mixing the phage and bacterial fractions, plating on standard solid media and checking at regular intervals for the presence of plaques (Hyman & Abedon, 2009). This process can be labour intensive and, given the non-specific nature of the technique, leaves the isolation of interesting phages to chance.

The Namib Desert (Figure 2.1) was selected as the environment to validate the reverse metaviromics concept due to the extreme nature, and the recent interest in the microbial communities it harbours (Prestel *et al*., 2008; Henschel & Lancaster, 2013; Ramond *et al*., 2014; Frossard *et al*., 2015). The Namib Desert, located on the southwestern tip of Africa, is estimated to be over 80 million years old, making it probably the oldest desert on earth (Burke, 2001). It covers an area of 130,000 km² with the Great Escarpment forming its inland boundary (Viles & Goudie, 2013). The majority of the Namib Desert ranges from a semiarid to a hyperarid environment (Hutchinson, 1995). While temperatures can be relatively stable along the coastline ranging between 9-20^{o}C, the central Namib is known to experience wide fluctuations, with a recorded range of below 0^{o}C to above 50^{o}C (von Willert, 1992; Makhalanyane *et al*., 2013). Rainfall has been reported to gradually increase from the coast

inland, averaging about 18mm annually (Makhalanyane *et al*., 2013). However, the desert is known to experience consecutive years with no rainfall, for example, the Swakopmund area once recorded a 10 year period without rainfall (Lancaster *et al*., 1983).

Studies have shown annual fog deposition to be more reliable than rainfall with a coefficient of variation (CV) of approximately 41% compared to 133% for rainfall (Seely & Henschel, 2000). Fog events are therefore considered to be the major source of water availability in the region (Büdel *et al*., 2009). These events originate offshore and are a result of the collision between the warm air from the Hadley Cell and the cold Benguela Current, and they extend approximately 60 km inland from the coast (Eckardt *et al*., 2013).

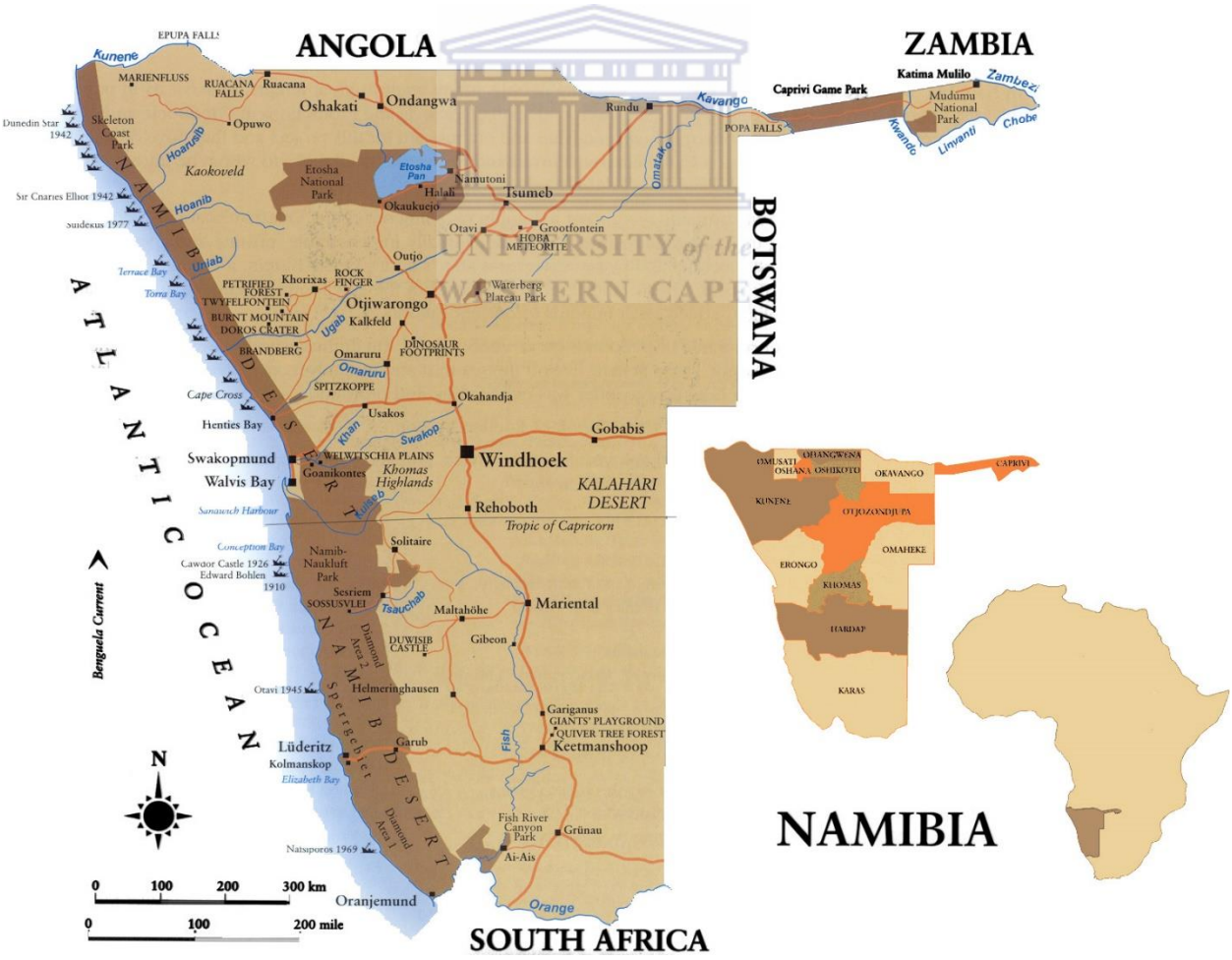

Figure 2.1: A map of Namibia showing the Namib Desert stretching across the coast inland. Image courtesy of The Fisheries Management, Namibia (2008).

The microbial community of the Namib desert has been understudied. An exploratory study was conducted by sampling surface sand from three different locations in the Namib desert (Prestel *et al*., 2008). The study used a combination of electron microscopy, 16S rRNA insert library and a small viral DNA linker amplification shotgun library. It showed that the bacterial communities are dominated by Firmicutes, especially those belonging to the genus *Bacillus* while other observed phyla included Planctomycetes, Bacteroidetes, δ-Proteobacteria and Chloroflexi (Prestel *et al*., 2008). The viral fraction showed mostly *Bacillus* associated siphoviruses (Prestel *et al*., 2008). This was the first microbial study conducted from the Namibian desert which only provided limited understanding of the true microbial diversity that exists in this environment.

More recent studies of the microbial community present in the Namibian desert have looked at the microbial ecology of hot and cold desert edaphic communities (Makhalanyane, 2012), the assembly of soil microbial and hypolithic communities along aridity gradients (Stomeo *et al*., 2013), the niche partitioning of edaphic bacterial communities in fairy circles found in the desert plains (Ramond *et al*., 2014) , the impact of wetting events on the response of bacterial community structure (Frossard *et al*., 2015), the habitat specific nature of edaphic bacterial communities found in the sand dunes (Ronca *et al*., 2015), and the functional capacity of soil niche communities (Vikram *et al*., 2015). These studies all used the culture-independent metagenomic approach to determine the microbial diversity present in the Namibian desert.

Similarly, a number of studies have applied the metagenomic approach to viral studies. Examples of such metavirome studies published from this region include an analysis of the Namibian hypolith phage communities (Adriaenssens *et al*., 2015). The study found that *Caudovirales* formed a significant proportion of phages found in the Namibian desert as previously observed (Prestel *et al*., 2008), but also suggested the possibility that unclassified viruses might be the dominant species. Another study highlighted the presence of novel

haloarchaeal salterproviruses and the high abundance of ssDNA viruses while investigating two hypersaline springs found in this region (Adriaenssens *et al*., 2016).

The common denominator with all these studies is that all the diversity analysis and community structure predictions were *in silico*, thus relying only on bioinformatic tools to understand the nature of the complex and diverse interactions within microbial communities and with the environment. This is particularly limiting for viral studies where it is very difficult to fully predict the interaction of a phage with its bacterial host based on genome data only. For example, while it is possible to predict which life cycle the phage undergoes based on the presence or absence of genes such as integrase or *cro*, it is impossible to estimate, from sequence data, important factors such as the number of phage progenies produced per bacterial cell, the time taken from cell infection to lysis or conditions under which the switch between lysogeny and lysis might be triggered. Therefore, a more informative overview is required, and this formed the basis of the study presented in this chapter.

Two sampling sites were selected in the Namibian desert: an abandoned copper prospecting site (Figure 2.2A) and a salt playa (Figure 2.2B). A physico-chemical analysis of the copper site showed the soil had a pH value of 7.3 and contained 1227.9 mg/kg of copper with other metals such as iron, manganese, arsenic, lead and zinc present in significant quantities (Hesse *et al*., 2016), a finding consistent with previously reported geomorphology and mineralogy analysis of the Namibian desert (Eckardt *et al*., 2013). Heavy metals such as copper have been shown to induce the lytic life cycle in temperate phages (Lee *et al*., 2006) and cause bacteriophage mortality, especially in the presence of water (Li & Dennehy, 2011). It would therefore be interesting to isolate phages from such a copper rich environment and attempt to understand any adaptive mechanism responsible for their persistence.

A B

Figure 2.2: A. Copper rich soil sample for the prospecting site B. Sample from the Swakop saline site consisting of microbial mat and spring water. Photos courtesy of Professor Ed Rybicki, University of Cape Town, South Africa.

Namibian salt playas are unique environments with a reported salinity range of 3-15% depending on the sampling depth, evaporation rate as a consequence of the time of the day, and the distance from the source (Day & Seely, 1988; Day, 1993; Eckardt *et al*., 2001; Eckardt & Drake, 2011). The Swakop saline site has a salinity range of 4.5%–8.6% (Adriaenssens *et al*., 2016). These values are relatively low in comparison to other saline environments such as the moderately hypersaline Salton Sea (CA, USA) which has a salinity range of 5% (water table) to 11.8% (sediments) (Swan *et al*., 2010) or crystallizer ponds and solar salterns with a range of 13.8% to 37% (Santos *et al*., 2010; Boujelben *et al*., 2012). This difference in salinity contributes to the novelty of this environment (Garcia-Heredia *et al*., 2012; Martínez-García *et al*., 2014). Physico-chemical analysis also showed that the Swakop saline site has a pH range of 6.5 (source) to 8.5 (sink), a conductivity range of 66–180 mS/cm and total dissolved solids (TDS) range of 42,000–115,000 ppm (Day & Seely, 1988; Eckardt & Drake, 2011; Adriaenssens *et al*., 2016). Temperatures have also been reported to reach 50^{o}C (Day & Seely,

1988). The combination of a temperature and salinity gradient suggests it could be an environment that harbours unique halotolerant and halophilic phages and bacterial hosts.

2.2 MATERIALS AND METHODS

2.2.1 Sample collection and preparation for microbial isolation

For the copper site, soil and rock samples were collected from a heap of copper containing material located at an abandoned copper prospecting site, approximately 5 km from the Gobabeb Desert Research and Training Station (23°33'35.21"S; 15°16'50.63"E) Namibia in April, 2013. The sample was re-suspended in 50 litres of distilled water. Swakop Saline (SS) samples were collected from the Eisfield salt playa 21.5 km North East of Swakopmund (22°29'5.31"S; 14°34'17.88"E). The samples consisted of 50 litres of stream water and sediment containing the microbial mat (Figure 2.2B) in an attempt to capture the overall diversity of the salt playa.

Samples from each site were initially allowed to stand for 2 hours with periodic shaking to loosen viruses from particulate matter and left to settle for a further 2 hours. The supernatant was decanted off and processed through a Millipore 1µm Polygard-CR filter (CR0101006) using a Millipore peristaltic pump (XX80EL230) to get rid of soil and other particles.

2.2.1.1 Phage purification

The 1µm filtrate was filtered through a 0.22µm Opticap® XL10 Durapore® filter (KVGLA10HH1) and the filtrate collected in a clean carboy. The 0.22µm filtrate was then subjected to tangential flow filtration (TFF) using a Millipore TFF cartridge holder (XX42PS001) and 30kDa cut-off filter (CDUF001TT). The approximately 50L of 0.22 µm filtrate was concentrated a final volume of about 200 ml. The TFF phage fraction was stored at 4°C pending further analysis.

2.2.1.2 Bacterial isolation

For the copper site, the TFF filter was backwashed using distilled water by reversing the Millipore TFF cartridge holder (XX42PS001) and the bacterial fraction collected in grenier tubes. For the Swakop saline site, fifty millilitres of the microbial mat was resuspended in spring water from the salt playa. Both samples were stored at 4^0C until required.

2.2.2 Metavirome sequencing and bioinformatics

Metaviromic DNA was extracted using a modified version of a published protocol (Sambrook, E. F. Fritsch, *et al*., 1989). Briefly, 1 ml of the TFF phage fraction was transferred to a 15 ml Sterillin tube and 12.5 µl of $MgCl_2$ was added to stabilize the DNA. Five microliters of DNAse I (1 U/ul, Fermentas) and 1 µl of RNAse A (10 mg/ml, Fermentas) were added and the mixture incubated at 37^{o}C for one hour to digest contaminating DNA and RNA. The absence of contaminating DNA was confirmed by PCR amplification of the 16S rRNA gene using the universal primers E9F (5'-GAGTTTGATCCTGGCTCAG-3') (Farrelly *et al*., 1995) and U1510R (5'- GGTTACCTTGTTACGACTT-3') (Reysenbach & Pace, 1995). Proteinase K (final concentration 97.5 µg/ml, Fermentas), 40 µl of 0.5 M EDTA and 50 µl of 10% SDS were added and the mixture and incubated at 55^{o}C for 2 hours with periodic mixing by gentle inversion. Complete lysis of phage particles was confirmed by adding 1 µl of 10,000x SYBR Gold (Life Technologies) to 10 µl of the mixture and observing under a fluorescence microscope. Viral DNA was extracted by adding an equal amount of Phenol: Chloroform: Isoamyl alcohol (25:24:1), gently inverting the tubes and centrifuging (Eppendorf 5810R) for 5 minutes at 13000 rpm at room temperature. The top aqueous layer was carefully removed and transferred into a sterile 2 ml Eppendorf tube. DNA was precipitated overnight at 4^{o}C by adding 2 volumes of absolute ethanol and 1/10 volume of 3M NaOAc (pH 5.2), followed by pelleting at 13000 rpm for 10 minutes. The DNA pellet was resuspended in 50 µl of TE buffer.

The metaviromic DNA was further cleaned using the Qiagen Gel Extraction kit (Qiaex II, cat. no. 20021). Ten nanograms was used to perform Phi29 amplification (GE Healthcare GenomiPhi HY DNA amplification kit) using the manufacturer's recommendations. The amplified DNA was sequenced using the Illumina MiSeq kit at the Sequencing facility in the Institute for Microbial Biotechnology and Metagenomics, University of the Western Cape, South Africa. Library preparation included a 10% phiX V3 spike as per the manufacturer's instructions (Preparation guide, Part #15031942, May 2012 revision) with the Illumina Nextera XT library prep kit/MiSeq Reagent kit V2.

For rapidly identifying novel phages for the purpose of this study, a basic analysis was carried out as follows: the raw sequence data was analyzed using CLC Genomics version 6.5 (CLC, Denmark). Raw reads were reference assembled to the phiX V3 genome used in the sequencing reaction for quality control. The unmapped reads from the initial reference assembly were mapped to the human genome to remove contaminating sequences. A *de novo* assembly was carried out on the remaining unmapped reads using the default CLC parameters (contig length- 200, mismatch cost- 2, insertion cost- 3, deletions cost- 3, length fraction- 0.5 and similarity fraction- 0.8). The unmapped reads used for *de novo* assembly were also uploaded to MetaVir online analysis program (Roux *et al*., 2014) which uses an array of bioinformatic resources including the GAAS tool (Angly *et al*., 2009), the NCBI and RefSeq databases, phylogenetic marker gene sequences from the pFAM database, and the alignment tool, MUSCLE (Edgar, 2004) to determine taxonomic composition and phylogenetic diversity.

Contigs of interest from the *de novo* assembly were individually uploaded to the phage search tool PHAST (Zhou *et al*., 2011) for identification of phage genes and annotation. Only open reading frames identified by both PHAST and CLC, using default parameters, were accepted and annotated.

2.2.3 Isolation and identification of saline bacterial isolates

The microbial mat resuspended in spring water was shaken at 50 rpm for 2 hours to encourage the dissociation of bacterial cells from debris. After a low speed centrifugation (500 rpm, 10 minutes), 10 µl of the supernatant was spread on agar plates of each medium type (Table 2.1) and incubated at 37°C until colonies were observed. Colonies of varying morphologies were selected after visual inspection and Gram staining. Each isolate was purified by three cycles of re-streaking on fresh agar plates and stored in broth of the medium of isolation supplemented with 50% glycerol at -80ºC until required.

Table 2.1: Media composition for isolating bacteria from the Swakopmund saline site.

Component	**Medium Composition (g/L), pH 8**					
	EHA	EHB	MHA	MHB	Medium A	Marine Agar
NaCl	250	125	80	98	200	19.45
KCl	2	-	-	2	1	0.55
K_2S0_4	-	5	-	-	-	-
$MgSO_4$. $7H_20$	20	-	20	1	20	-
$MgCl_2$. $6H_20$	-	50			-	8.8
$CaCl_2$. $6H_20$	-	0.12	-	0.36	-	1.8
$FeCl_3$	0.023	-	-	0.001	-	0.1
Na- Citrate	3	-	3	-	-	-
Casamino acid	7.5	-	7.5	-	0.5	-
Yeast Extract	10	5	1	10	0.5	1
Tryptone/ Peptone	-	5	5	5	0.5	5
NaBr	-	-	-	0.23	-	-
KBr						0.08

$NaHCO_3$	-	-	-	0.06	-	0.16
$Fe(NH_4)_2(SO_4)_2.\ 6H_2O$	-	-	0.05	-	-	-
K_2HPO_4	-	-	0.5	-	0.3	-
Glucose	-	-	-	1	0.5	-
Starch	-	-	-	-	0.5	-
Sodium pyruvate	-	-	-	-	0.3	-
Sodium sulphate	-	-	-	-	-	3.24
Trace Minerals	-	-	-	Yes	-	Yes
Bacteriological Agar	20	20	20	20	20	20

2.2.3.1 16S rRNA identification of saline site (SS) bacterial isolates

The 16S rRNA gene from each isolate was amplified using the universal primers E9F (5'-GAGTTTGATCCTGGCTCAG-3') (Farrelly *et al*., 1995) and U1510R (5'-GGTTACCTTGTTACGACTT-3') (Reysenbach & Pace, 1995). DNA templates were prepared by re-suspending a loop-full of each isolate in 100 µl of PCR grade water and boiling at 85ºC for 5 minutes. Each PCR reaction contained 12.5 µl of KAPA Taq ReadyMix, 0.5mM final primer concentration and 2 µl of DNA template and adjusted to a final volume of 25 µl with water. The PCR conditions were: initial denaturation at 95ºC for 3 min, followed by 35 cycles of denaturation at 95ºC for 30s, annealing at 55ºC for 30s and extension at 72ºC for 2 min, with a final extension at 72ºC for 2 min. The PCR products were analyzed for the correct size by gel electrophoresis (section 2.2.3.2), followed by purification using the Nucleospin® Gel and PCR Clean-up kit (Macherey-Nagel) according to the manufacturer's instruction. The amplicons were sequenced using E9F and U1510R primers on an ABI PRISM® 377 automated DNA sequencer at the Central Analytical Facility of the University of Stellenbosch, South Africa.

2.2.3.2 Gel electrophoresis

PCR products were analyzed on a 1 - 2 % [w/v] agarose gel depending on the size of the expected fragments. Gels were prepared using 1X TAE buffer and contained ethidium bromide at a final concentration of 0.5 µg/ml. DNA samples were prepared by the addition of 6X DNA loading buffer prior to loading onto the agarose gel. Electrophoresis was performed in 1x TAE running buffer at 10-12 Volts/cm for 1 hour and DNA size estimated using lambda DNA digested with *Pst*I. DNA was visualized and imaged using a UV (302 nm) Transilluminator (Alpha Imager, Alpha Innotech, USA).

2.2.4 Phage-host screening

2.2.4.1 Thalassomonas

A PCR reaction was performed to confirm the authenticity of the assembly a novel phage identified from the bioinformatic analysis of the copper site metavirome. The reaction was performed in a thermal cycler Gene Amp PCR system targeting a 5kb region using primers TH5FOR (5'- CTCTCGACCCTGCCTTTCAA - 3') & Th5REV (5' - CCCATACAGATAGCTGAACT - 3') and 200 ng of DNA template. The 50 µl reaction was carried out in 0.2 ml thin walled tubes containing 1X Phusion buffer, 200 µM each of dNTPs, 0.5 µM of each primer, 200 ng of template DNA and 0.02 U/µl of Phusion DNA polymerase (Thermo Scientific). The following conditions were used: 1 cycle of 98°C for 3 minutes; 34 cycles of 98°C for 10 seconds, 58°C for 30 seconds, and 72°C for 2.5 minutes; and 72°C for 10 minutes; and refrigeration at 4°C.

Potential bacterial hosts for the novel phage identified from the copper site metavirome assembly were investigated by challenging known species of *Thalassomonas* with the environmental phage fraction and PCR screening. Some of the *Thalassomonas* species have since been reclassified as *Thalassotalea*. *Thalassotalea loyana* (ID: 280483), *Thalassomonas viridans* XOM25 (ID: 137584), *Thalassotalea agariperforans* (ID: 864068), *Thalassotalea ganghwensis* (ID: 221989), *Thalassotalea agarivorans* (ID: 349064), *Thalassomonas actiniarum* (ID: 485447) and *Thalassomonas haliotis* (ID: 485448) were obtained from the ATCC, CCUG and DSMZ culture collections. Each strain was assayed for phage infection using the standard soft agar overlay technique (Adams, 1959). Briefly, each strain was used to inoculate 10 ml of marine broth (Difco) and grown overnight at 200 rpm and at the recommended temperature for each strain (Macián *et al*., 2001; Efrony *et al*., 2009; Hosoya *et al*., 2009; Park *et al*., 2011). Five hundred microlitres of the overnight cultures were used to inoculate fresh 50 ml marine broth and grown until each culture reached $O.D._{600}$ 0.4. Two

hundred microlitres of these cultures were added to 1.5 ml tubes containing 100 µl of serially diluted environmental phage fraction and incubated at room temperature for 10 minutes. The mixture was added to 3 ml marine broth containing 0.5% bacteriological agar and spread evenly on marine agar plates. The plates were checked for the presence of plaques every 24 hours.

2.2.4.2 Saline spring

Saline site bacterial isolates were similarly evaluated as described for *Thalassomonas* species in Section 2.2.4.1, using the media of isolation (Table 2.1), as potential hosts for a novel phage.

2.3 RESULTS AND DISCUSSION

2.3.1 Metavirome assembly

Detailed descriptions of the bioinformatic analysis of the metagenomes of the copper and saline sites have been presented elsewhere (Adriaenssens *et al*., 2016; Hesse *et al*., 2016). The copper site was found to harbour diverse viral species from the order Caudovirales and several unclassified viruses. The ssDNA circovirus-like species were found to be dominant in this environment possibly due to the bias introduced through the multiple displacement amplification employed. Four novel genomes were identified including three small circovirus-like genomes and a *Thalassomonas* phage BA3-like viral genome that will be discussed later in this Chapter. Sequences representative of the families *Circoviridae, Microviridae* and *Inoviridae* were found to represent a significant portion of the viral diversity of the Swakop saline site. The authors cited proximity of the site to the sea and the potentially frequent visits of a number of bird species as possible explanations for the dominance of these viral species. The metaviromic analysis also revealed homologs to almost all known haloarcheal viral genomes, suggesting the possibility of novel archaea-infecting phages.

A sequence assembly and bioinformatic analysis was conducted parallel to the published/submitted studies described above to rapidly identify novel phage targets for this study. A total of 941, 600 raw reads were generated for the copper site metavirome with an average paired read length of 246 bp representing 255,102,403 bp. Reference assembly to PhiX and the human genome using CLC removed 4,760,946 bp suggesting minimal human contamination. The *de novo* assembly generated 20,721 contigs, ranging from 173 to 41,006 bp in length. Using the online analysis tool, MetaVir, 38,499 predicted genes and 125 circular contigs were identified. Only 3519 contigs showed significant similarity to known sequences on the NCBI database, while no matches were found for 17,202 contigs, further confirming the novelty of this environment. Sequencing the Swakop saline metaviromic DNA generated 4,060,839 raw reads with an average paired read length of 371.04 bp, after reference assembly to PhiX and the human genome. The *de novo* assembly using CLC generated 22,282 contigs. The contigs ranged in size from 297 to 23854 bp. MetaVir identified 37,605 predicted genes and 73 circular contigs. About a quarter of the contigs (4558) were similar to known sequences, leaving 17724 novel sequences.

2.3.2 COPPER SITE

2.3.2.1 Viral diversity and taxonomic composition

The copper site metavirome was dominated by sequences of viruses belonging to the order *Caudovirales* (67%), the order *Phycodnaviridae* represents 7% and includes *Chloroviruses* and *Prasinoviruses*, while a further 4% are classified as *Mimiviridae*. The remaining 22% represent unclassified phages. No RNA viruses were detected because the nucleic acid extraction method biased for DNA viruses.

The major phage families are represented with siphoviruses as the dominant family (40%), followed by myoviruses and podoviruses (29% and 24% respectively) while the rest (7%) are unclassified (Figure 2.3).

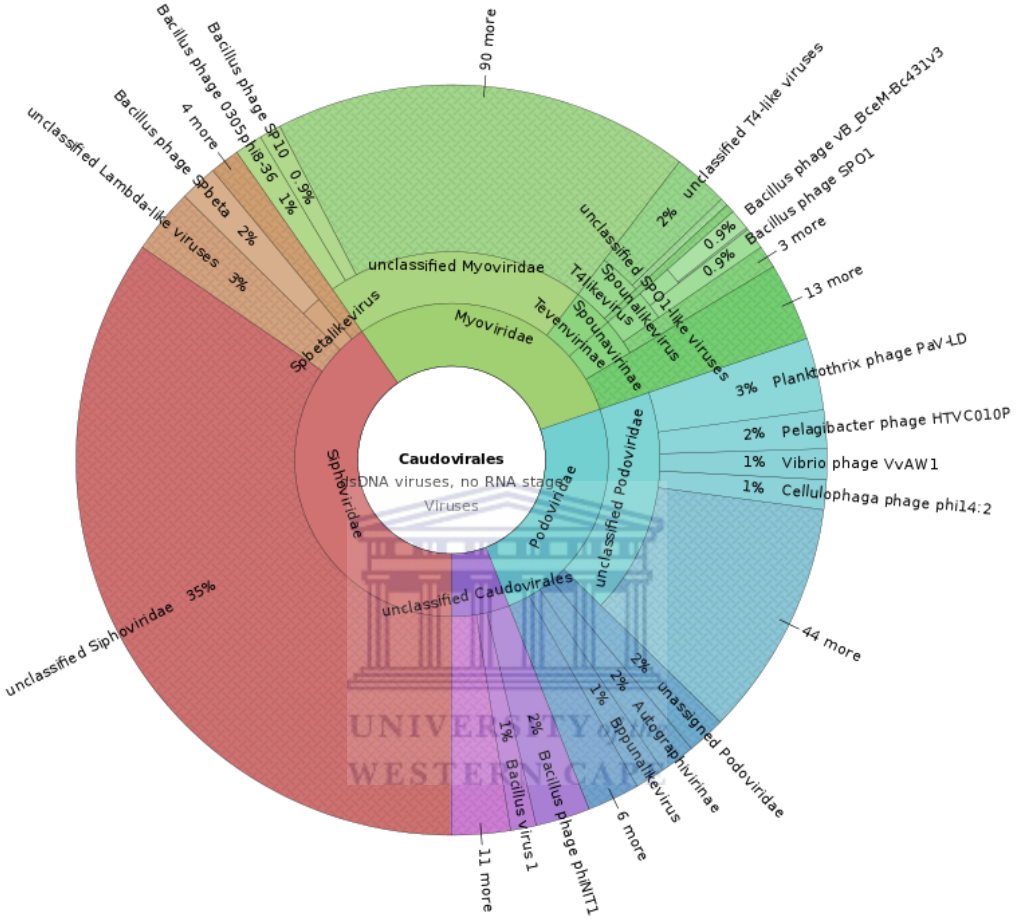

Figure 2.3: Graphical representation of the phage diversity of the copper site using the online sequence data analysis tool, Metavir (Roux *et al*., 2014). The environment is dominated by Caudovirales, forming 67% of the phage community.

Bacillus related phages were found to be the most dominant species in the copper site metavirome. A phage similar to *Bacillus* phage Spbeta was the dominant siphovirus representing 4% of all siphoviruses; and a *Bacillus* phage 0305phi8-36-like phage was the most dominant myovirus (4%). A few podoviruses were more represented than others e.g. *Planktothrix* phage PaV-LD (13%), *Pelagibacter* phage HTVC010P (7%), *Vibrio* phage

VvAW1 and *Cellulophaga* phage phi14:2 (both 5%) and *Thalassomonas* phage BA3 (4%) (Figure 2.4).

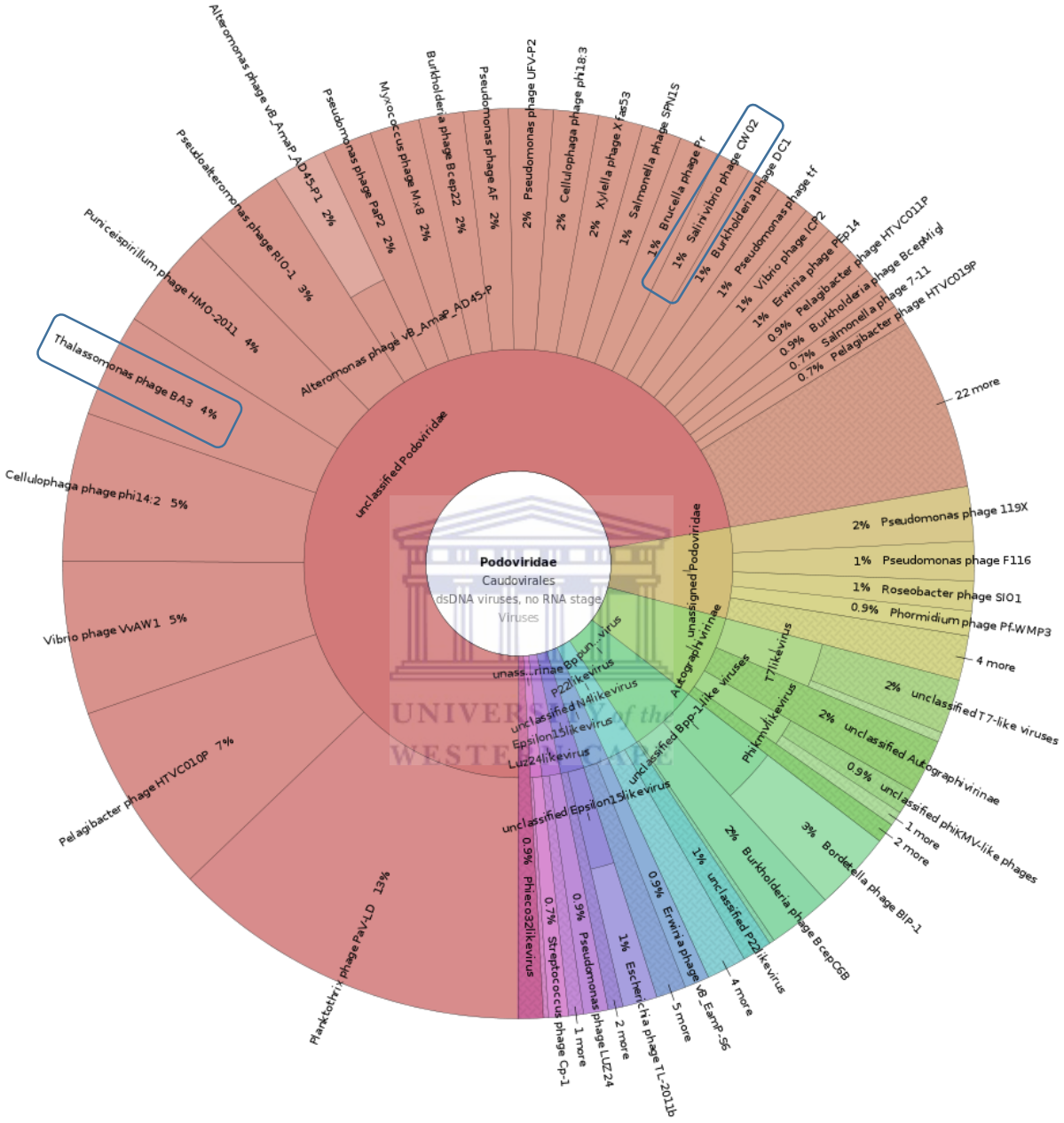

Figure 2.4: Graphical representation of one of the phage families (podoviruses) found in the copper site using the online sequence data analysis tool, Metavir (Roux *et al*. 2014). The most represented podoviruses include *Planktothrix* phage PaV-LD (13%), *Pelagibacter* phage HTVC010P (7%), *Vibrio* phage VvAW1 and *Cellulophaga* phage phi14:2 (both 5%) and *Thalassomonas* phage BA3 (4%).

2.3.2.2 Identification of novel phages

A summary of the closest phage hits for the largest assembled contigs are provided in Table 2.2. Most of the assembled contigs had no significant similarity to phages on the NCBI database. However, one contig of particular interest was contig 13 which was the most covered (1156.49x) and showed high similarity to phage BA3 (72% nucleotide identity over 49% of the genome). The coverage could have been overestimated due to at least one reason. Phi29 DNA polymerase used to increase the amount of DNA available for sequencing uses random-primed DNA amplification based on isothermal multiple strand displacement and has been shown to introduce copy number biases (Arriola *et al*., 2007). However, given that the phage was not the only whole genome in the metavirome but was still the most covered, it is reasonable to assume that it was one of the dominant species in the copper environment, at least at the time of sampling. This could possibly have been confirmed using qPCR.

Table 2.2: A summary of the largest contigs assembled from the copper site metavirome sequencing using CLC Genomics version 6.5 (CLC, Denmark) with default parameters.

CONTIG	AVERAGE COVERAGE	CLOSEST BLASTn PHAGE HIT	QUERY COVER (%)	E-VALUE	IDENTITY (%)	ACCESSION NUMBER
13 (35 kb)	1156.49	*Thalassomonas* phage BA3	49	0.0	72	EU124666.1
71 (26kb)	110.54	*Ralstonia* phage RSL1	0	0.79	82	AB366653.2
237 (24kb)	18.73	*Cronobacter* phage CR8	0	2e-15	87	KC954774.1
21 (23kb)	64.75	*Pseudomonas* phage PaBG	0	0.0058	81	KF147891.1
693 (22kb)	25.91	*Enterobacteria* phage Vb EcoM	0	2.4	85	KP007362.1

Phage BA3 has previously been shown to be an effective therapeutic agent in treating Coral white plague disease caused by *T. loyana* (Efrony *et al*., 2009). Furthermore, BA3 is the only known phage that infects the genus *Thalassomonas*. Considering the Copper site is distant from any marine environment, it begged the question as to why a phage related to one that infects a marine microorganism should be present, especially so abundantly, in this sample. A hypothesis was that the host might be a terrestrial *Thalassomonas* species in which case it would be the first report, seeing all currently known *Thalassomonas* are of marine origin. The second explanation was that if the BA3-like phage infected another bacterial species or even genus, the study of the phage-host interaction would improve our knowledge of the evolution and relatedness of phages. It also posed an interesting question about how phages are spread

and whether or not they can survive in an environment devoid of the host bacterium if transferred there by other means (wind, rain/fog, animals). For all these reasons, the research here focused on isolating and characterizing this novel phage-host pair.

2.3.2.3 Analysis of the BA3-like phage genome

The contig is 35461 bp in size, and has a GC content of 40.44%. The genome of the BA3-like phage (designated Contig13) is circularly permuted and contains 47 open reading frames (ORFs). ORFs 1-39 oriented counterclockwise, while ORFs 40-47 clockwise (Table 2.3). ORFs 40 and 42 were found to be homologous to genes involved in DNA binding and replication (early genes). Five ORFs (2, 10, 14 and 17, 18) are similar to late genes such as rRNA adenine demethylase, stabilization protein, portal protein and the large and small terminase subunits respectively. Two genes potentially involved in host lysis were found (ORFs 1 and 21) and the genome also contains a structural gene (ORF 12), possibly a coat protein. Fourteen Contig13 genes were similar to those found in phage BA3 (between 36-82% similarity) but no specific functions could be assigned to them.

Table 2.3: Table summarizing predicted ORFs of the Contig13 phage based on BLASTx hits on the NCBI database. It also highlights genes that are homologous to those found in phage BA3. The + or – sign before the gene name is indicative of the the strand on which the gene is located.

ORF NUMBER	ORF LENGTH	START	STOP	% IDENTITY	% COVERAGE	E-VALUE	SCORE	HIT DEFINITION	DESCRIPTION IN PHAGE BA3
1/-	1749	1	1750	67.34	28	1.00E-80	279	Hypothetical protein BA3_0001 [*Thalassomonas* phage BA3]	host lysis gene
2/-	5705	1753	7458	81.31	97	0	1932	Hypothetical protein BA3_0002 [*Thalassomonas* phage BA3]	late gene: rRNA Adenine dimethylase
3/-	839	7502	8341	61.97	100	3.00E-122	362	Hypothetical protein BA3_0005 [*Thalassomonas* phage BA3]	-
4/-	767	8352	9119	63.03	93	5.00E-93	285	Hypothetical protein BA3_0006 [*Thalassomonas* phage BA3]	-
5/-	563	9107	9670	82.97	97	6.00E-110	323	Hypothetical protein BA3_0007 [*Thalassomonas* phage BA3]	-
6/-	341	9680	10021	36.28	100	7.00E-11	64.7	Hypothetical protein [*Sphingobium chungbukense*]	

7/-	842	10026	10868	71.15	19	9.00E-18	90.5	Hypothetical protein BA3_0008 [*Thalassomonas* phage BA3]	-
8/-	1376	10868	12244	75.11	100	0	740	Hypothetical protein BA3_0009 [*Thalassomonas* phage BA3]	late gene: similar to stabilization protein
9/-	749	12247	12996	75.9	100	2.00E-128	375	Hypothetical protein BA3_0010 [*Thalassomonas* phage BA3]	-
10/-	335	13001	13336	42.86	88	4.00E-19	86.3	Hypothetical protein similar to phage protein (N4 Gp49/phage Sf6 gene 66) family [*Pseudomonas* phage KPP25]	other virus
11/-	599	13346	13945	51.22	98	9.00E-60	196	Hypothetical protein BA3_0011 [*Thalassomonas* phage BA3]	-
12/-	1247	14026	15273	86.51	100	0	752	Hypothetical protein BA3_0013 [*Thalassomonas* phage BA3]	structural gene: similar to coat protein

13/-	923	15289	16212	56.83	100	4.00E-104	318	Hypothetical protein BA3_0014 [*Thalassomonas* phage BA3]	-
14/-	2144	16297	18441	75.21	99	0	1122	Hypothetical protein BA3_0015 [*Thalassomonas* phage BA3]	late gene: similar to portal protein
15/-	125	18434	18559	-	-	-	-	-	
16/-	200	18546	18746	43.94	100	6.00E-09	56.2	Hypothetical protein BA3_0016 [*Thalassomonas* phage BA3]	-
17-	1640	18743	20383	80.22	100	0	965	Hypothetical protein BA3_0017 [*Thalassomonas* phage BA3]	late gene: similar to terminase large subunit
18/-	653	20425	21078	57.07	94	4.00E-77	242	terminase small subunit [*Vibrio* phage VvAW1]	
19/-	146	21068	21214	66.67	88	3.00E-09	56.6	Hypothetical protein BA3_0020 [*Thalassomonas* phage BA3]	similar to gp26 *Burkholderia cenocepacia* phage BcepB1A; YP_024873

20/-	509	21214	21723	36.36	98	1.00E-34	129	Hypothetical protein BA3_0021 [*Thalassomonas* phage BA3]	-
21/-	404	21735	22139	58.27	95	5.00E-46	156	Hypothetical protein BA3_0022 [*Thalassomonas* phage BA3]	similar to host lysis gene
22/-	575	22172	22747	40.43	98	8.00E-40	145	Hypothetical protein [*Marinobacter subterrani*]	
23/-	416	22744	23160	44.07	43	7.00E-06	51.2	Hypothetical protein [*Vibrio coralliilyticus*]	
24/-	260	23168	23428	-	-	-	-	-	
25/-	350	23425	23775	53.33	52	3.00E-11	65.1	Hypothetical protein [*Shewanella baltica*]	
26/-	425	23763	24188	50.81	88	6.00E-34	125	Hypothetical protein [*Vibrio shiloi*]	
27/-	218	24239	24457	39.68	88	3.00E-07	53.1	Hypothetical protein [*Lactobacillus vaginalis*]	
28/-	371	24457	24828	50	85	8.00E-29	113	Hypothetical protein BA3_0027 [*Thalassomonas* phage BA3]	-

29/-	287	24941	25228	52.63	60	3.00E-09	60.1	Hypothetical protein [*Shewanella* sp, ZOR0012]	
30/-	569	25225	25794	-	-	-	-	-	
31/-	287	25791	26078	41.27	65	2.00E-04	45.4	Hypothetical protein VPAG_00041 [*Vibrio* phage douglas 12A4]	other virus
32/-	518	26075	26593	-	-	-	-	-	
33/-	251	26580	26831	-	-	-	-	-	
34/-	821	26966	27787	61.99	99	7.00E-118	350	Hypothetical protein BA3_0030 [*Thalassomonas* phage BA3]	-
35/-	383	27851	28234	73.39	98	3.00E-60	192	Hypothetical protein BA3_0031 [*Thalassomonas* phage BA3]	-
36/-	293	28291	28584	-	-	-	-	-	
37/-	1151	28640	29791	39.11	71	3.00E-54	193	Hypothetical protein [*Pseudoalteromonas piscicida*]	
38/-	950	29795	30745	46.86	96	7.00E-84	267	Hypothetical protein [*Pseudoalteromonas piscicida*]	

39/-	683	30742	31425	48.12	94	3.00E-60	201	Pentapeptide repeat protein [*Leptospira interrogans*]	
40/+	227	31554	31781	57.97	92	5.00E-21	88.6	Hypothetical protein BA3_0035 [*Thalassomonas* phage BA3]	early gene: DNA binding
41/+	812	31784	32596	46.42	93	1.00E-57	196	Hypothetical protein BA3_0036 [*Thalassomonas* phage BA3]	-
42/+	698	32598	33296	43.22	50	5.00E-23	103	DNA replication protein [*Serratia odorifera*]	
43/+	224	33289	33513	45.16	84	3.00E-09	58.2	Hypothetical protein [*Endozoicomonas montiporae*]	
44/+	161	33576	33737	-	-	-	-	-	
45/+	464	33737	34201	34.51	73	5.00E-14	74.7	Hypothetical protein [*Providencia alcalifaciens*]	
46/+	230	34198	34428	50	87	2.00E-11	64.3	Hypothetical protein [*Pseudoalteromonas piscicida*]	
47/+	674	34415	35089	44.89	79	1.00E-36	137	Hypothetical protein [*Tolumonas* sp, BRL6-1]	

2.3.2.4 Isolation of the BA3-like phage (Contig13)

A 5kb region of Contig13 phage genome was amplified to authenticate the contig. Successful amplification and end sequencing of the region of the Contig13 phage from the copper site metaviromic DNA showed that the phage genomic DNA was present in the original environmental phage fraction. This ruled out the possibility of a pseudo-genome (assembly artifact), at least over the area amplified. Seven *Thalassomonas* species obtained from culture collections were exposed to the environmental phage fraction, however, no plaques were detected on agar plates. There could be 2 explanations: i) none of the *Thalassomonas* species screened could serve as the host, or ii) that the phage was present in the sample but not viable. The fact that the phage fraction was DNAse treated before genomic extraction suggests that the phage genome was at least within an intact capsid. The structure required for attachment to the bacteria, for example the tail fibre, may have been damaged.

2.3.2.5 Could a *Thalassomonas* species be the host?

Phages generally bind to receptors such as proteins, teichoic acids, lipopolysaccharides or flagella on the cell surface of the host (Riede *et al*., 1985; Rakhuba *et al*., 2010; Chaturongakul & Ounjai, 2014). Phage structures such as the tail fibre and base plate have been shown to play some role in host recognition and the mechanical injection of genetic material (Fiers *et al*., 1976; Mayer, 2007). Contig13 phage encodes a glycoprotein (gp49, ORF 10) with 42.8% identity to a cell surface antigen recognition protein in *Pseudomonas* phage KPP25 (Arm *et al*., 1991). No similar protein was found on the genome of phage BA3, and while not conclusive, suggests that these phages might target different receptors on the cell membrane of the bacterial host or indicate that they infect different bacterial species.

Phage genomes are known to be arranged as functional modules which can be exchanged with other prophages or infecting phages through recombination events, leading to new combinations of modules which can result in potentially new and viable phages (Botstein, 1980; Hatfull, 2008). Similarities in the genes and arrangement of modules can therefore be indicative of evolutionary relationships between phages and hosts (Lima-Mendez *et al*., 2011). Comparative analysis of the phage BA3 and the Contig13 phage showed that the suite of genes involved in each phase of the life cycles of these phages are markedly different (Figure 2.5).

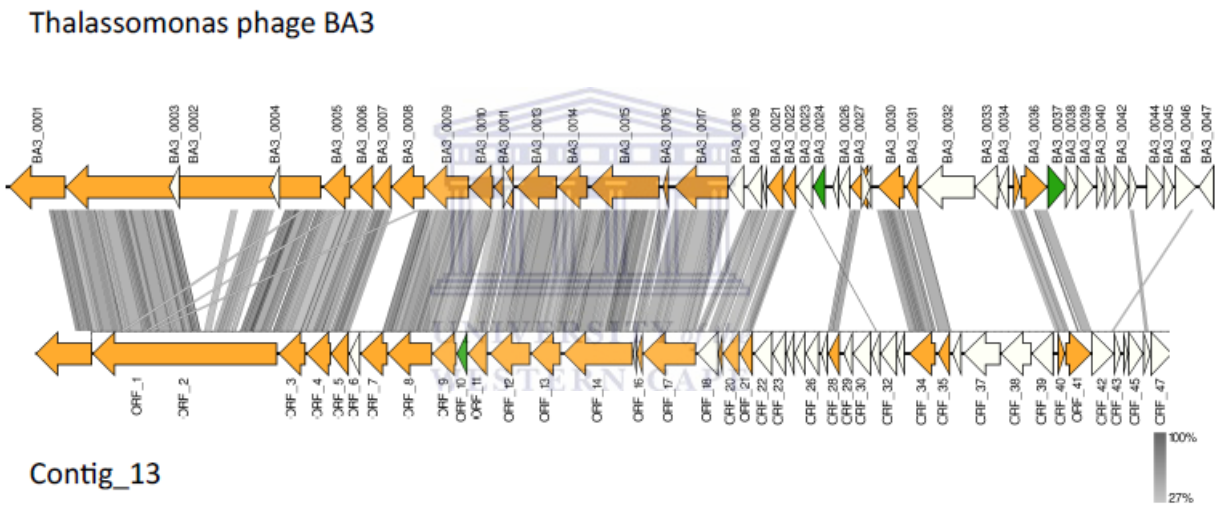

Figure 2.5: Comparative analysis of phage BA3 and Contig13 identified from the Namibian metavirome using Easyfig genome comparison visualizer (Sullivan *et al*., 2011). The white arrows represent genes unique to each genome, orange arrows represent conserved genes and green arrows represent genes that are potentially involved in host recognition.

For example, there are seven recognizable early genes (ORFs 26, 32, 33, 35, 39, 44, 45) in the genome of phage BA3 but only one (ORF 35) is similar to any of the genes on Contig13 phage (ORF 40, 58% similarity, 92% query cover, E-value 2 e-29). Such differences suggest that the host for the Contig13 phage might not necessarily be *T. loyana* or possibly not even a *Thalassomonas* species. Given that phages tend not to carry unnecessary genes, it is reasonable

to assume that the metabolic requirements, hence susceptible bacterial hosts, of these phages might be different (Canchaya *et al*., 2003; Lang *et al*., 2012).

However, while 20 of the 47 predicted ORFs in BA3 have no similarity to any of the genes on the Contig13 phage, 11 of these are most similar to marine bacteria or phages. For example, ORF 25 is most similar to a hypothetical protein from *Shewanella baltica*, a bacterium first isolated from the Baltic sea and which has been shown to be a major cause of spoilage in marine fish stored at low temperature (Fonnesbech Vogel *et al*., 2005). Other examples include ORFs 37, 38, and 46 which all show similarity to *Pseudoalteromonas piscicidia*, a bacterium associated with egg disease in damselfish (Nelson & Ghiorse, 2002). Perhaps the most telling example is ORF 26 which is most similar to a hypothetical protein of *Vibrio shiloi*, a bacterium implicated in coral bleaching disease (Banin *et al*., 2000). As stated earlier, phage BA3 infects *Thalassomonas loyana* which itself causes the coral white plague disease, (Efrony *et al*., 2009). Interestingly, *V. shiloi* was isolated from corals in the Mediterranean Sea and *T. loyana* from corals from the Red Sea, two water bodies directly connected by the Suez Canal. This is the strongest indication that even if Contig13 phage does not infect a *Thalassomonas* species, it could be of marine origin.

Given the inability to identify the host for the Contig13 phage, the question remains as to how a phage, possibly of marine origin, is relatively abundant in the desert gravel plains. A possible explanation could be the frequent occurrence of fog events in the Namibian desert. Indeed, a number of studies have found that bacteria form a significant part of these fog events, with approximately 20% of all particles in the range of 0.25-1 μm diameter being viable bacterial cells (Bauer *et al*., 2003; DeLeon-Rodriguez *et al*., 2013). It has also been suggested that atmospheric samples consistently contain up to 17 bacterial taxa (DeLeon-Rodriguez *et al*., 2013). Another study of microbial aerosols detected 31 bacterial genera with 5 dominant genera (*Vibrio, Bacillus, Pseudoalteromonas, Psychrobacter, Salinibacterium*) (Dueker *et al*., 2012).

It is therefore possible that the marine species found in the copper site were transported there by fog events. Phages, being smaller, can potentially be transported the same way. If this was the case for the BA3-like phage, it will be interesting to study how much impact fog events have on the microbial diversity of the Namibian desert.

Another possible reason is human interference. The sampling site was close to an abandoned copper mine which would have required large volumes of water for processing. The only realistic source of water would be the sea. It is possible that the phage and bacterial host were transported inland while supplying the prospect site with water. However, given that prospecting activities ceased over 60 years ago, the phage would have had to persist in the soil for decades. While the persistence of phages in the environment for such an extended period of time has not been reported, phage PRD1 has been shown to persist and remain infective after 22 years when it was stored in broth at 4°C (Ackermann *et al*., 2004). While possible, it is unlikely that the BA3-like phage would have survived that long.

Another likely explanation for finding a marine phage in the Namibian desert could be attributed to the above average rainfall recorded in many parts of sub-Saharan Africa -including Tanzania, Madagascar, Cameroon, Congo and Namibia- in 2010 and 2011. Multi-satellite Precipitation Analysis (MPA) showed that in early May 2011, more rainfall was recorded in one day than was typical of an entire year. More specifically, 70 millimetres of rain was recorded compared to less than the averaged 20 millimetres per year (NASA Earth Observatory, 2011).

The high precipitation resulted in the formation of a supercell thunderstorm (Figure 2.6), which caused large volumes of water to be deposited in areas along the Namibian seashore and desert plains. This event could have caused a significant, if temporal, change in the microbial diversity

of the Namibian desert, introducing species that were not native to the region. The host of the BA3-like phage or the phage itself could have been one of these species.

Figure 2.6: Image of a 'shelf cloud' formed over the Central Namibian desert. A strong downdraft causes the cloud to spread out at the ground level, producing an advancing wedge of condensed water which is dumped quickly over a small area. Photograph courtesy of Dr. Frank Eckardt, University of Cape Town, South Africa.

In summary, these are hypotheses and would require further investigation such as an attempt to understand the impact of fog events and above average rainfall on the microbial diversity of the region over time.

The failure to identify the host for the Contig13 phage using bacterial strains from culture collections prompted a modification of the research approach. The new approach involved

targeting bacteria from the sampling site, using a combination of metaviromic data, physico-chemical analysis and the predicted bacterial host from BLAST analysis, culturing the bacteria and challenging a lawn of each isolate with the environmental phage fraction. This could not be tested for the Contig13 phage due to lack of samples, hence the decision to investigate the Swakop saline site.

2.3.3 SWAKOP SALINE SITE

2.3.3.1 Viral diversity and taxonomic composition

The Swakop saline site was also dominated by *Caudovirales* (66%) as expected. Other orders that were represented include *Phycodnaviridae* (6%), *Mimiviridae* (4%), and single stranded DNA viruses such as *Microviridae* (5%), *Circoviridae* (2%). Myoviruses were the dominant family accounting for 50% of all *Caudovirales* compared to 33% of siphoviruses and 14% of podoviruses (Figure 2.7).

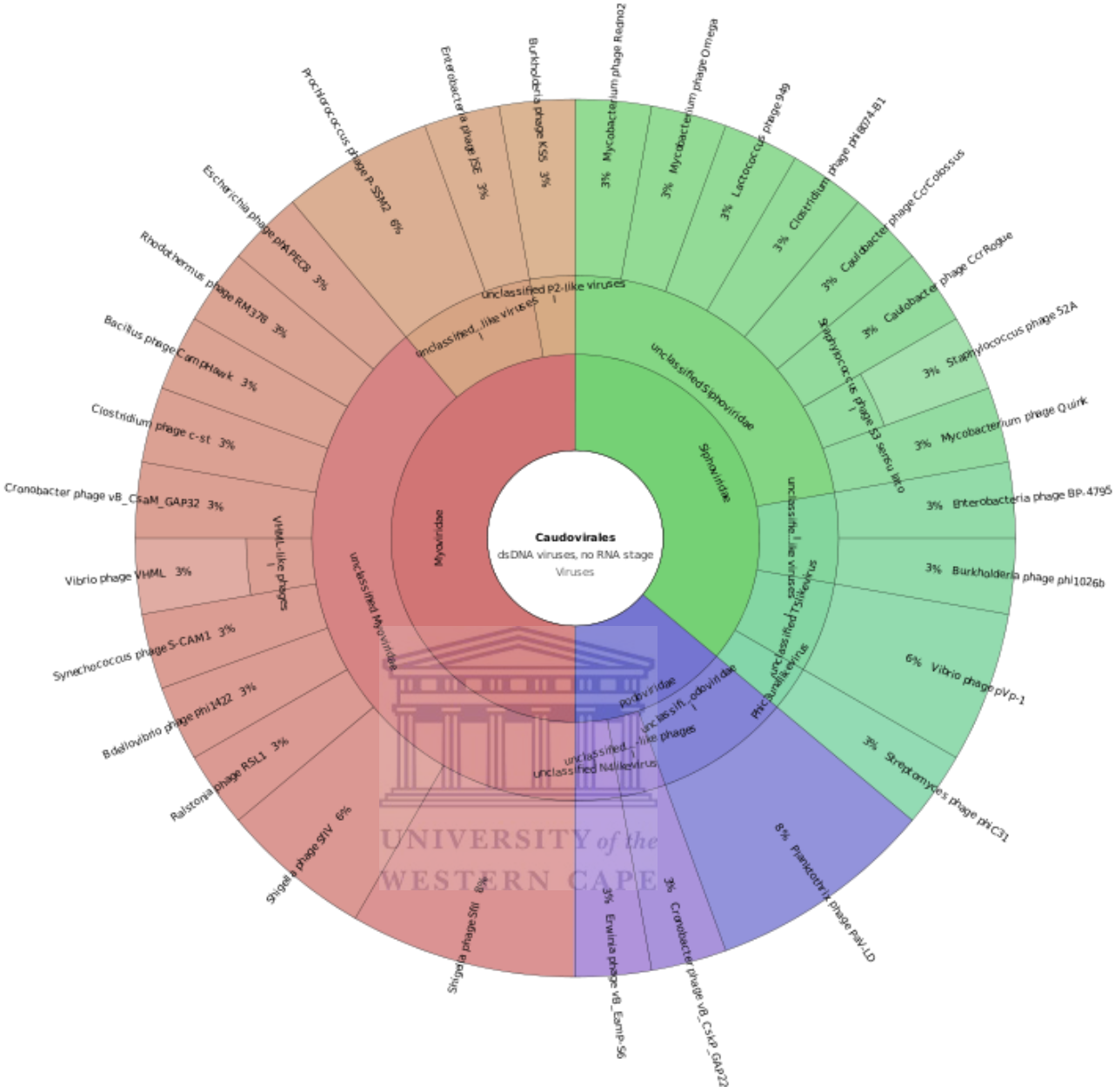

Figure 2.7: Graphical representation of the *Caudovirales* diversity of the saline site using the online sequence data analysis tool, MetaVir (Roux *et al*. 2014). The environment is dominated by Caudovirales, forming 66% of the phage community.

The most dominant viral species was *Bacillus* related, similar to the copper site metavirome with the myovirus *Bacillus* phage G representing 7% of all *Caudovirales*. This supports the observation that *Bacillus* species are generally more abundant in saline environments compared to other environments (Horikoshi, 1999) and are the most commonly isolated bacteria in

haloalkaline lakes (Mwirichia *et al*., 2010). No particular sipho- or podovirus was found to be dominant, however, *Salinivibrio* phage CW02-like viruses were found to represent 2% of all podoviruses (Figure 2.8).

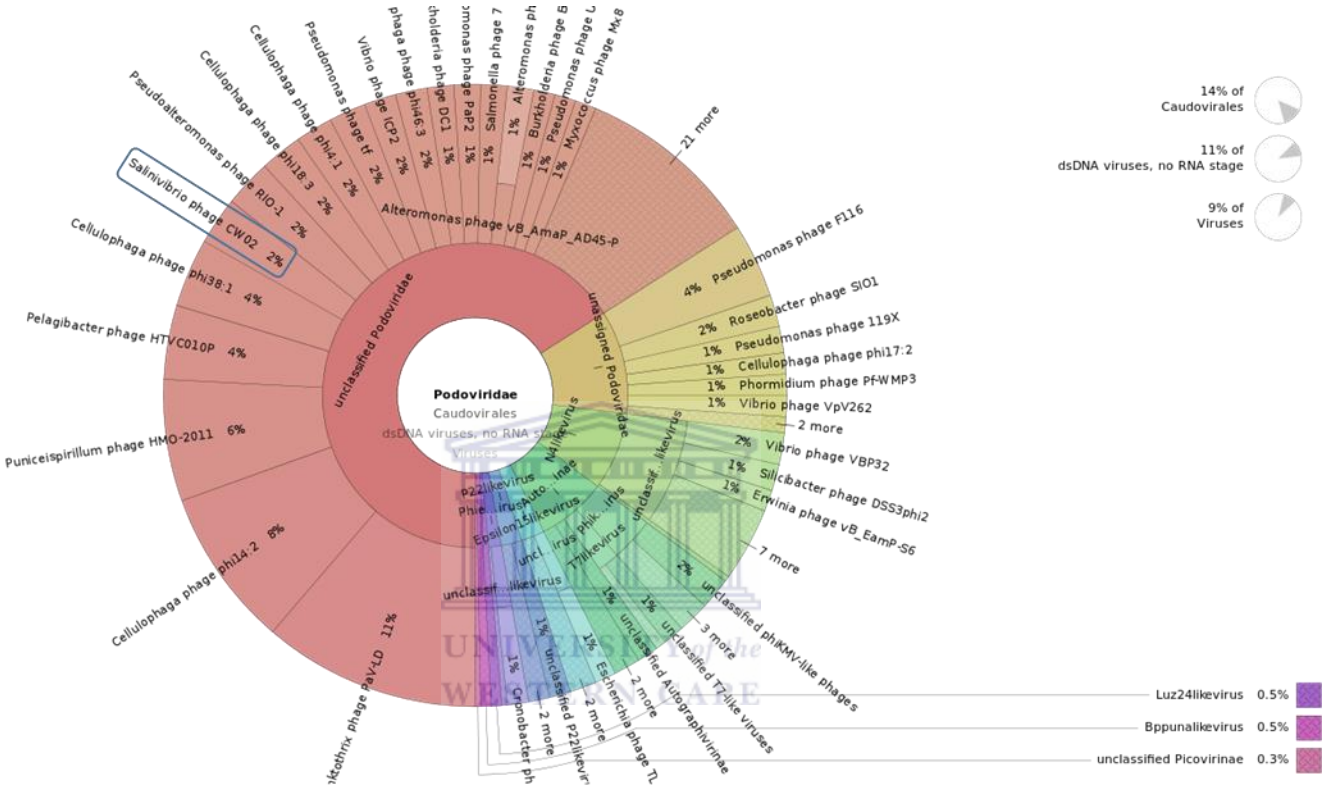

Figure 2.8: Graphical representation of the podoviruses found in the saline site using the online sequence data analysis tool, Metavir (Roux *et al*. 2014). The most represented podoviruses include *Planktothrix* phage PaV-LD (11%), *Cellulophaga* phage phi14:2 (8%), and *Puniceispirillum* phage HMO-2011 (6%).

2.3.3.2 Identification of novel phages

A summary of the BLAST analysis of some of the largest contigs assembled from the Swakop saline site is provided in Table 2.4. Most of the contigs had no significant similarity to known phages, as reported for the copper site, further affirming the novelty of these environments.

Table 2.4: A summary of the largest contigs assembled from the Swakop saline site metavirome sequencing using CLC Genomics version 6.5 (CLC, Denmark) with default parameters.

CONTIG	AVERAGE COVERAGE	CLOSEST BLASTn PHAGE HIT	QUERY COVER (%)	E-VALUE	IDENTITY (%)	ACCESSION NUMBER
481 (22kb)	24.54	*Mycobacterium* phage Cosmo	0	0.016	76%	KP027195.1
179 (20kb)	67.30	*Dactylococcopsis salina* PCC 8305 prophage	7	5 E 83	75	CP003944.1
622 (20kb)	28.11	*Bacillus* phage Vb _BpuM_BpSp	0	3.2	90	KT895374.1
534 (15kb)	16.86	Uncultured Mediterranean phage uvMED	0	0.14	82	AP013370.1
1007 (13kb)	14.90	*Thermoanaerobacterium* phage THSA-485A	0	0.001	85	CP003186.1

However, of particular interest was a potential *Salinivibrio* infecting phage. First, viral diversity and taxonomic composition analysis showed that *Salinivibrio* infecting phages were present at both the copper (Figure 2.4) and saline (Figure 2.8) sampling sites. A reference assembly of the Swakop saline site contigs to the copper site reads identified at least 4 contigs with similarity to the only characterized *Salinivibrio* infecting phage CW02 (Table 2.5). The similarity was also low enough to suggest that the phage was novel.

Table 2.5: A summary of the Swakop saline contigs showing similarity to the only characterized *Salinivibrio* phage CW02

CONTIG	CLOSEST BLASTn PHAGE HIT	QUERY COVER (%)	E-VALUE	IDENTITY (%)	ACCESSION NUMBER
3968 (842 bp)	*Salinivibrio* phage CW02	35	4e-6	65	JQ446452.1
6086 (691bp)		12	3E07	75	
14311 (423 bp)		39	8e-5	68	
11113 (347 bp)		19	4e-8	81	

While not the most abundant phage species, this research focused on isolating a novel *Salinivibrio* infecting phage for a number of reasons. First, *Salinivibrio* infecting phages were present at both sampling sites as stated earlier. Given the marked differences between these sites, especially in terms of salt concentrations, there might be the possibility that the phages play a role in shaping the physiology of the host to survive such harsh conditions (Clokie *et al*., 2011). Second, there are only two known *Salinivibrio* infecting phages, CW02 and UTAK (Shen *et al*., 2012) therefore it would be interesting to isolate more novel phages infecting this understudied genus and evaluate their biotechnological potential. Third, apart from 16S rRNA sequence data for some of the species, very little is known about this genus, therefore a study of the interaction of phages infecting these potential hosts will contribute to the body of knowledge of these bacterial species.

2.3.3.3 Bacterial isolation

Previously characterized *Salinivibrio* species were shown to be moderate halophiles which grew optimally at 37ºC, an ideal isolation temperature given that the saline site is known to reach 50ºC (Day & Seely, 1988; Huang *et al*., 2000). These species are also known to tolerate NaCl concentrations up to 12.5% which is higher than the reported range of 4.5-8.6% for the saline site (Huang *et al*., 2000; Shen *et al*., 2012; Adriaenssens *et al*., 2016). An optimum pH of 7.3 and the presence of glucose and yeast were also reported to be ideal for these species because they are capable of both aerobic and anaerobic metabolism (Huang *et al*., 2000; Shen *et al*., 2012). These characteristics were taken into consideration during media design, hence two halophile-targeting media (MHA and MHB, Table 2.1) were included during bacteria isolation. The two media had similar NaCl concentrations (8% and 9.8% respectively) but varied in terms of the presence, absence and concentration of other salts.

Eight bacterial strains were isolated from the Swakop saline site. Each isolate has been discussed based on the top BLASTn hit to their 16s rRNA gene on the NCBI database:

SS1 was most similar to *Idiomarina loihiensis*, a γ-proteobacterium, first isolated from a hydrothermal vent at a depth of 1,300m on the Loihi submarine volcano, Hawaii and relies on amino acid metabolism for carbon and energy rather than sugar fermentation. It has also been shown to survive a wide range of temperatures (4 – 46ºC) and salinities (0.5 -20% NaCl) (Hou *et al*., 2004).

Two (SS2 and SS3) *Salinivibrio* species were isolated. SS2 was most similar to *Salinivibrio costicola subsp. vallismortis strain DV,* a halotolerant facultative anaerobe initially isolated from a hypersaline pond located in Death Valley, California (Huang *et al*., 2000). SS3 is most similar to *Salinivibrio costicola* BNH, first isolated from the largest hypersaline playa in Iran and is a possible producer of hydrolytic enzymes (Unpublished data).

Table 2.6: Summary of the bacteria species isolated from the Swakop saline site.

Isolate	Isolation Media	16S rRNA Sequence length (bp)	Closest Hit	Query Cover (%)	E-value	Identity (%)	Accession Number
SS1	MHA	1326	*Idiomarina loihiensis* strain L2TR	99	0.0	96	NR074933.1
SS2	Marine Agar	1311	*Salinivibrio costicola* subsp. vallismortis strain DV	97	0.0	95	NR028703.1
SS3	MHB	1431	*Salinivibrio costicola* strain BNH	99	0.0	96	KC594859.1
SS4	Medium A	1352	*Halomonas caseinilytica* strain NY-7	100	0.0	99	JN903901.1
SS5	Marine Agar	1397	*Marinobacter xestospongiae* strain UST090418-1611	100	0.0	99	NR109066.1
SS6	Medium A	1336	*Halomonas sinaiensis* strain AloSharm	100	0.0	96	NR115003.1
SS7	EHB	1426	*Virgibacillus salarius* strain SA-Vb1	99	0.0.	99	NR041270.1
SS8	EHB	825	*Halomonas eurihalina* strain ATCC 49336	100	0.0	95	NR026250.1

SS5, *Marinobacter xestospongiae*, a Gram-negative, rod-shaped and slightly halophilic bacterial strain, was isolated from the marine sponge *Xestospongia testudinaria* collected from the Red Sea coast of Saudi Arabia. It tolerates a maximum of 6% (w/v) NaCl, and optimal growth occurred at 2.0% (w/v) NaCl, pH 7.0-8.0 and 28-36 °C. When grown on marine agar, colonies are cream-coloured, circular and slightly irregular, convex with a smooth surface and an entire edge, semi-transparent (Lee *et al*., 2012).

SS7, *Virgibacillus salarius* strain SA-Vb1, moderately halophilic bacterium was first isolated from a salt-crust sample collected from Gharsa Salt Lake (Chott el Gharsa), Tunisia. It is quite robust and grows over a wide temperature range (10-50^{0}C) but optimally at 30- 35^{0}C, and 0.5-25% NaCl (optimally at 7-10%) and pH 5.5 -10 (optimally at 7.5) (Hua *et al*., 2008).

There is a dearth of information about *Halomonas caseinilytica* (SS4), *Halomonas sinaiensis* (SS6) and *Halomonas eurihalina* (SS8) except that this study found them to grow at high salinity (20-30%), pH 8 and at 37^{0}C.

Apart from the *Salinivibrio* species, none of the other isolates have any described phages. Due to the low similarity to known species (≤ 96% identity), SS1, 2, 3, 6, and 7 are likely to be novel bacterial species which can be further evaluated, not only for phage-host relationships, but for their biotechnological potential.

2.3.3.4 Isolation of a novel phage

The two *Salinivibrio* species isolated were challenged with the saline site phage fraction using the soft agar overlay technique (Adams, 1959). Plaques were detected only on agar plates for isolate SS3. Extending the incubation time of other culture to 76 hours did not yield further plaques.

2.4 CONCLUSION

This study has revealed the viral diversity of two understudied sites in the Namibian desert using metaviromic tools. The usefulness of incorporating metaviromic sequence data in the process of identifying novel phages was also demonstrated and attempts were made to isolate two novel phages, the Contig13 phage and a *Salinivibrio* infecting phage. While only the *Salinivibrio* phage was successfully isolated, the analysis of Contig13 showed the potential impact of natural events on shaping viral diversity. The *Salinivibrio* phage is described in detail in the next chapter.

3

CHARACTERIZATION OF PHAGE

SMHB1

3.1 INTRODUCTION

As presented in Chapter two, using metaviromic sequence data as a guide, two possibly novel *Salinivibrio* species were isolated, and upon infection with the environmental phage preparation, plaques were observed on the lawn of SS3. This chapter therefore focuses on the isolation and characterization of the SS3-infecting phage, SMHB1. Phages CW02 and UTAK are the only phages known to infect *Salinivibrio* species, however, only phage CW02 has been described in terms of general features and genomics while the host for UTAK may have been incorrectly reported (Goel *et al*. 1996; Shen *et al*. 2012).

Phage SMHB1 infects a halobacterium, which based on 16S rRNA analysis has been identified as *Salinivibrio costicola*, isolated from a salt playa in the Namibian desert. The isolate was designated as SS3. There are seven known *Salinivibrio* species (based on 16S rRNA data) on the NCBI database and three have been whole genome sequenced (as of 4th April, 2016). In-depth studies of these species have not been conducted, but it is thought that the possession of genes involved in DNA repair, resistance to harsh environmental conditions and heavy metals and possibly, light based energy production, are important factors that enable them to thrive in extreme conditions (Gorriti *et al*., 2014). Therefore, understanding the interaction between these species and phages that infect them may reveal interesting co-evolutionary mechanisms that contribute to the aforementioned capabilities of these bacteria. Such relationships could also be explored for biotechnological applications such as in the development of *in vitro* packaging kits for fosmid library construction and the use of phage derived products like polymerases and lysozymes for industrial and research purposes.

3.2 MATERIALS AND METHODS

3.2.1 Isolation of pure phage stock

A single, isolated plaque was excised from a plate using a cut tip, resuspended in 200 µl of distilled water and left overnight at 4ºC to encourage phage dissociation from the agar plug. The suspension was centrifuged at 9000 rpm for 5 minutes. The supernatant was removed and filtered through a 0.22 µm filter prior to plaque assay. This process was repeated twice. Five plaques were excised from the final plate, used to infect an actively growing culture of the bacterial host, SS3 and incubated overnight at 37ºC with shaking at 200 rpm. The culture was centrifuged at 9000 rpm for 10 minutes and the supernatant was filtered through a 0.22 µm filter. The phage lysate was stored in 25% glycerol at -80ºC.

3.2.2 Transmission electron microscopy (TEM)

Phage samples were prepared for imaging by centrifugation of the phage lysate (Centrifuge 5417 R, Merck) at 16000 rpm for 5 hours and the phage pellet resuspended in 100mM ammonium acetate, pH 6.5 (Ackermann, 2009). Three microliters of the phage suspension was spotted onto carbon coated 200 mesh copper grids and allowed to adsorb for 1 minute. The carbon grid was washed twice in sterile water and stained with 2% aqueous uranyl acetate for 30 seconds. The samples were viewed using a LEO 912 Omega TEM (Zeiss, Oberkochen, Germany) at 120 kV and images collected using a ProScan CCD camera.

3.2.3 Host range testing

A number of bacteria were challenged with the phage to determine the host range. Bacterial isolates from the saline site included two *Salinivibrio* strains, SS2 and the host for phage SMHB1, SS3. The other isolates were: *Halomonas caseinilytica, Halomonas eurihalina, Halomonas sinaiensis, Idiomarina loihiensis*, *Marinobacter xestospongiae*, *Virgibacillus salarius*. The isolates were tested as described in Section 2.2.4.1.

3.2.4 Extraction of genomic DNA for phage SMHB1 and *Salinivibrio* isolates

A soft agar layer overlay assay (Section 2.2.4.1) was carried out using SMHB1 phage concentrations that resulted in plates with plaques that were too numerous to count (TNTC plates). The plates were flooded with MHB broth (Chapter 2, Table 2.1) and kept at 4^{o}C overnight. The flooded soft agar layer was carefully scraped into sterile 50 ml falcon tubes and centrifuged at 8000 rpm for 10 minutes. The supernatant was filtered through a 0.22 μm filter to remove bacterial cells and stored at 4^{o}C until required.

An overnight culture of the bacterial host, SS3, was inoculated into a fresh one litre of MHB broth (1% v/v) and grown until O.D.$_{600}$ 0.4 was reached. One millilitre of the phage lysate was added and the culture was incubated at 37^{o}C with shaking for 48 hours. Cell debris was removed by centrifugation at 9500 rpm for 30 minutes. The supernatant was further purified by passing it through a 0.22 μm filter. Phage particles were concentrated by adding PEG 8000 to a final concentration of 7% and incubating overnight at 4^{o}C. Viral particles were pelleted at 9500 rpm for 15 minutes and resuspended in SM buffer and DNA was extracted as described in the previous Chapter (Section 2.2.2).

Genomic DNA was also extracted from the two *Salinivibrio* species (SS2 and SS3) isolated from the saline site. Five millilitres of overnight cultures of each strain in MHB broth was centrifuged at 5000 rpm for 5 minutes. Genomic DNA was extracted from the bacterial pellets using a modified protocol (Sambrook, E. F. Fritsch, *et al*., 1989). Briefly, each pellet was resuspended in 950 µl of TE buffer and 50 µl of 10% SDS and 5 µl of 20 mg/ml proteinase K (Fermentas) were added. The suspension was incubated at 37ºC for 1 hour. One hundred and eighty microlitres of 5M NaCl and 150 µl of CTAB/NaCl solution were added, mixed thoroughly and incubated at 65ºC for 20 minutes. An equal volume of Chloroform/ Isoamyl alcohol was added. The solution was mixed gently and centrifuged at 14000 rpm for 10 minutes at room temperature to separate the phases. The DNA-containing supernatant was transferred to a fresh tube and extracted with an equal volume of phenol/ chloroform/ isoamyl alcohol, mixed gently and centrifuged at 14000 rpm for 10 minutes at room temperature. The clear upper phase was transferred to a new tube, 0.6 volume isopropanol was added, gently mixed and centrifuged at 14000 rpm for 10 minutes at room temperature. The pellet was washed with 70% ethanol, air dried and resuspended in 200 µl of 1X TE buffer.

3.2.5 Preparation of clone library of phage SMHB1 genomic DNA

3.2.5.1 SMHB1 genomic DNA restriction digest

The phage genomic DNA was digested with six restriction enzymes (*Eco*RI, *Hind*III, *Swa*I, *Hha*I, *Mbo*I, *Rsa*I) according to the manufacturer's instructions (Fermentas). The 20 µl reaction mixture contained 1µg of DNA, 2 µl of 2X reaction buffer and 1U of restriction endonuclease. The reaction was incubated for 6 hours in a water bath at 37ºC (30ºC for *Swa*I). The digestion product was analyzed by electrophoresis on a 1% agarose gel as described in Section 2.2.3.2.

3.2.5.2 Blunt-end cloning

Based on the restriction pattern observed, the product of *Rsa*I digest was cleaned using the Nucleospin® Gel and PCR Clean-up kit (Macherey-Nagel) and cloned into pJET 1.2/blunt vector (Life Technologies) according to the manufacturer's instructions. The reaction was performed in a thermal cycler Gene Amp PCR system. The ligation reaction was prepared on ice and contained the following: 10 µl of 2X reaction buffer, 150 ng DNA, 1 µl of pJET 1.2/blunt cloning vector (50ng/ µl), 1 µl of T4 DNA ligase (5U/ µl). The volume was made up to 20 µl using nuclease-free water. The reaction was incubated at 22^{o}C for 30 minutes and used directly for transformation.

3.2.5.3 Preparation of electrocompetent cells

A single colony of *E. coli* Epi300 was inoculated into 10 ml of Luria broth and incubated overnight at 37^{o}C with shaking at 250 rpm. One millilitre of the overnight culture was used to inoculate 100 ml of Luria broth and incubated at 37^{o}C with shaking at 250 rpm until OD_{600} 0.5. The culture was cooled on ice for 15 minutes with periodic swirling and centrifuged at 2500 rpm for 20 minutes at 4^{o}C. The pellet was washed three times with 10 ml of ice cold distilled water with centrifugation at 2500 rpm for 20 minutes at 4^{o}C. After washing, the pellet was gently resuspended in 4 ml of ice cold 10% glycerol and centrifuged at 2500 rpm for 20 minutes at 4^{o}C. The supernatant was discarded and the pellet was resuspended in 1 ml 10% glycerol. Fifty microlitres of the cell suspension was aliquoted into pre-cooled 1.5 ml microfuge tubes and stored at -80^{o}C.

3.2.5.4 Transformation

An aliquot (50 µl) of electrocompetent *E. coli* EPI 300 cells was thawed on ice for 5 minutes. Five microlitres of the ligation reaction was added and the mixture was transferred into pre-cooled electroporation cuvettes. Electroporation was carried out using a Genepulse™ electroporator (Bio-Rad Laboratories, Hercules, CA, USA) at 25 µF capacitance, 1.8 kV and 200 ohms resistance. One millilitre of Luria broth was immediately added to each cuvette after electroporation. The transformed cells were transferred into sterile microfuge tubes and incubated at 37ºC for 1 hour. Fifty microlitres of the transformation mixture was plated out onto LB agar supplemented with ampicillin (200 µg/ml) and incubated at 37ºC for 16 hours.

3.2.5.5 Plasmid DNA extraction using the Alkaline lysis method

Transformants were inoculated into 5 ml of Luria broth supplemented with ampicillin (200 µg/ml) and incubated overnight at 37ºC with shaking at 250 rpm. Two millilitres of the cultures was centrifuged at 5000 rpm for 5 minutes. The pellets were resuspended in 110 µl of resuspension buffer containing 50 mM glucose, 25 mM Tris-HCl, pH 8, 10 mM EDTA and 1 µl of RNAse A (Fermentas). Two hundred microlitres of the lysis buffer (200 mM NaOH, 0.5% SDS) was added. The suspension was mixed by gentle inversion and incubated at room temperature for 5 minutes. One hundred and fifty microlitres of the neutralization buffer (11.5% glacial acetic acid, 3M potassium acetate) was added and mixed by gentle inversion. The tubes were incubated on ice for 5 minutes followed by centrifugation at 13000 rpm for 10 minutes at room temperature. Three hundred microlitres of the supernatant was transferred into sterile tubes and 600 µl of 100% ethanol was added. The mixture was centrifuged at 13000 rpm for 10 minutes at room temperature and the supernatant discarded. The DNA pellets were washed once with ice cold 70% ethanol and air dried. The pellets were resuspended in 100 µl of TE buffer.

3.2.5.6 Clone de-replication and sequencing

One microgram of plasmid DNA from positive clones was digested with *Swa*I (Fermentas) as described in Section 3.2.5.1 above. Insert sizes were analyzed by electrophoresis on a 1 % agarose gel as described in the previous Chapter (Section 2.2.3.2). Four amplicons were sequenced using the pJET forward primer (5'-CGACTCACTATAGGGAGAGCGGC-3') on an ABI PRISM® 377 automated DNA sequencer at the Central Analytical Facility of the University of Stellenbosch, South Africa.

3.2.6 Whole genome sequencing and analysis

SMHB1, SS2 and SS3 genomic DNA were further cleaned using the Qiagen Gel Extraction kit (Qiaex II, cat. no. 20021) and sequenced as described in Section 2.2.2 of the previous Chapter. Analysis of the sequence data was carried out using CLC Genomics version 6.5 (CLC, Denmark). Raw reads were reference assembled to phiX V3 genome used in the sequencing reaction for quality control. Unmapped reads were mapped to the human genome to remove any contaminating sequences. A *de novo* assembly was carried out on the remaining unmapped reads using default parameters.

For phage SMHB1, the longest assembled contig was uploaded to the phage search tool PHASTER (Arndt *et al*., 2016) for identification of phage genes. Only open reading frames identified by both PHASTER and CLC Genomics version 6.5 (CLC, Denmark) were accepted and annotated.

Closely related phages were identified using PHASTER (Arndt *et al*., 2016) and RAST (Aziz *et al*., 2008). Direct repeats were identified using REPFIND (Betley *et al*., 2002) with a 10-bp minimum repeat length. tRNA genes were predicted using ARAGORN (Laslett & Canback, 2004) and the tRNAscan-SE program (Lowe & Eddy, 1997; Schattner *et al*., 2005). Intron

prediction was done using the RNAweasel server (Gautheret & Lambert, 2001). *In silico* restriction analysis was carried out using NEBcutter V2.0 (http://nc2.neb.com/NEBcutter2/index.php). Tetranucleotide usage analysis was carried out using TETRA (Teeling *et al*., 2004). Transmembrane regions were predicted using the TMHMM server v2.0 (Krogh *et al*., 2001). For phylogenetic tree construction, the full-length amino acid sequences of selected terminase proteins were aligned using MEGA6 (Tamura *et al*., 2013), and the tree was constructed using the built-in program (Saitou & Nei, 1987).

Ideograms were constructed using Circoletto online genome comparison tool (Darzentas, 2010) to compare the SMHB1 genome with other similar phage genomes identified using BLAST analysis. The other phage genomes were *Vibrio* phage PV94 (NC027368.1), *Vibrio* phage K139 (NC003313.1) and a prophage identified within the genome of *Salinivibrio* sp KP-1 (NZLAQR00000000.1).

For the genomes of SS2 and SS3, basic analysis of the assembled contigs were carried out using RAST server (Aziz *et al*., 2008), BLAST analysis against the NCBI database and CLC Genomics version 6.5 (CLC, Denmark) with default parameters. Clustered regularly interspaced short palindromic repeats (CRISPRs) regions were scanned for using CRISPRFinder (Grissa *et al*., 2007).

3.3 RESULTS AND DISCUSSION

3.3.1 TEM morphological characterization

The morphology of phage SMHB1 was studied using transmission electron microscopy and was identified as a member of the family *Myoviridae* with a head-neck-contractile tail geometry (Ackermann, 2007). The phage has a capsid diameter of approximately 56 nm which is within the range of 30 to 60 nm reported for 65% of aquatic viruses (Wommack & Colwell, 2000). Phage SMHB1 has a tail length of approximately 106 nm (Figure 3.1).

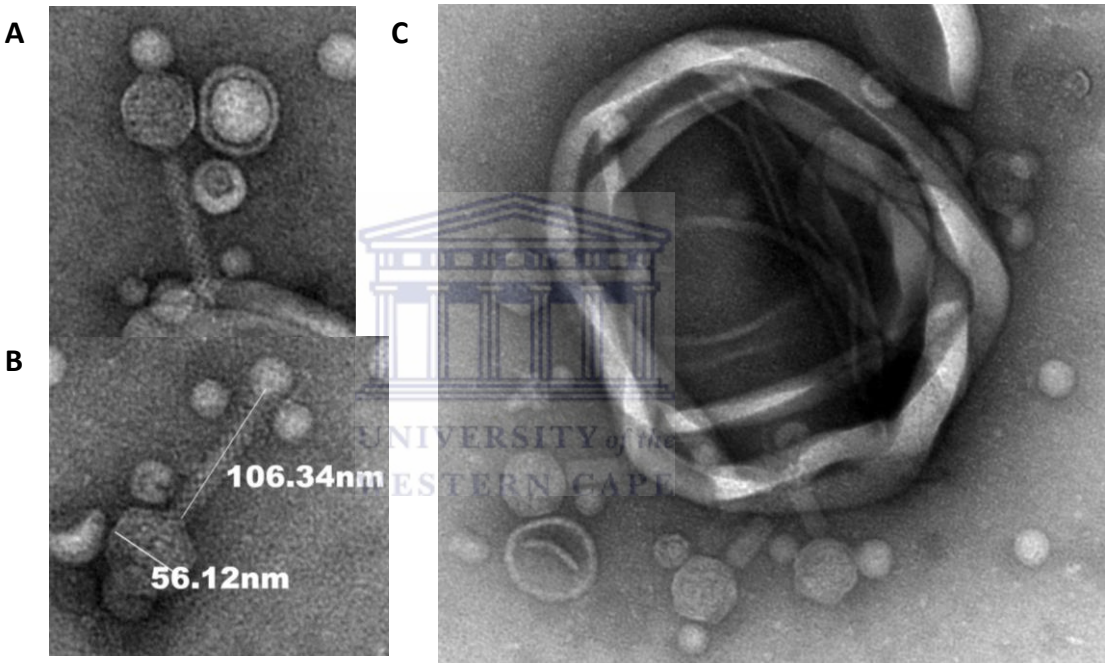

Figure 3.1: A & B. Electron micrographs of phage SMHB1 C. Electron micrograph showing phage attachment to cellular debris after cell disruption. The tails are compressed due to the injection of genetic material in a syringe-like manner.

3.3.2 Determination of phage host range

All eight saline site isolates were tested for susceptibility to phage SMHB1. This was done as it was recently shown that some phages are capable of infecting across several bacteria phyla (Malki *et al*., 2015). Four myoviruses from Lake Michigan infected *Pseudomonas aeruginosa*, *E. coli*, *Arthrobacter* sp., *Chryseobacterium* sp., and *Microbacterium* sp. (Malki *et al*., 2015),

and represents the first time that such a broad host-range has been identified. The phages demonstrated varied phenotypic characteristics despite similarities at the genomic level. The authors proposed that such a broad host range could be related to the oligotrophic nature of the environment, and this ability could offer the phages a competitive advantage. Given that the saline spring that SMHB1 was isolated from could also be considered oligotrophic, its ability to infect various hosts isolated from that same environment was tested. However, no other bacteria from the isolates tested was found to be susceptible to phage SMHB1, and although not an exhaustive screen, suggests that the host range may be limited to this particular strain as is the case for many other bacteriophages. Attempts were made to acquire other *Salinivibrio* species reported in literature. However, many are either not present in culture collections or could not be provided by the authors of manuscripts describing them. Future studies could investigate other *Vibrio* species as possible hosts.

3.3.3 Phage genomic DNA restriction digest analysis

Phage SMHB1 genomic DNA was digested with six restriction enzymes (Figure 3.2). Uncut or partially digested DNA is usually an indication of the presence of modifications of the phage DNA to avoid susceptibility to restriction endonucleases of the bacterial host (Samson *et al*., 2013; Vasu & Nagaraja, 2013). Such modifications usually involve the addition of a methyl or glycosyl group (Krüger & Bickle, 1983; Adams & Burdon, 1985).

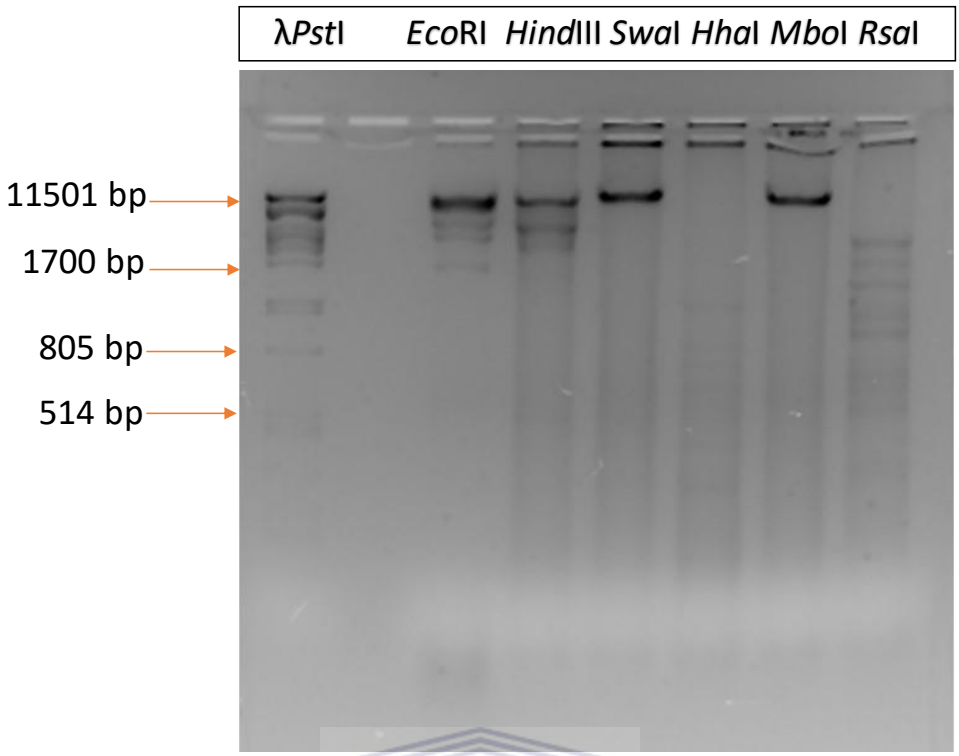

Figure 3.2: Agarose gel (1%) electrophoresis showing restriction patterns generated from digesting phage SMHB1 DNA with six restriction enzymes. Lane one, Lambda phage DNA cut with *Pst*I used as a molecular marker.

*Eco*RI, *Hind*III and *Rsa*I are insensitive to *dam* or *dcm* methylation (Gardner *et al*., 2014) and therefore were able to digest the phage DNA. *Mbo*I is sensitive to *dcm* methylation and the phage DNA remained uncut. This suggests methylation at the C5 position of the second cytosine in the sequences CCAGG and CCTGG within the genome of phage SMHB1 (Marinus & Morris, 1973; May & Hattman, 1975). Before going ahead with full genome sequencing, the small fragments generated from *Rsa*I digestion were used to prepare a small insert library to ensure that the phage isolated was not any of the already described phages.

3.3.4 Phage genomic DNA library construction

Plasmid DNA from transformants was digested with *Swa*I to determine insert size and for de-replication. Four clones, lanes 5, 9, 11 and 14, were selected and sequenced (Figure 3.3). The inserts were approximately 1.5 kb, 3kb, 2kb and 900 bp respectively. Sequencing in one direction generated about 700 bp of data for each clone.

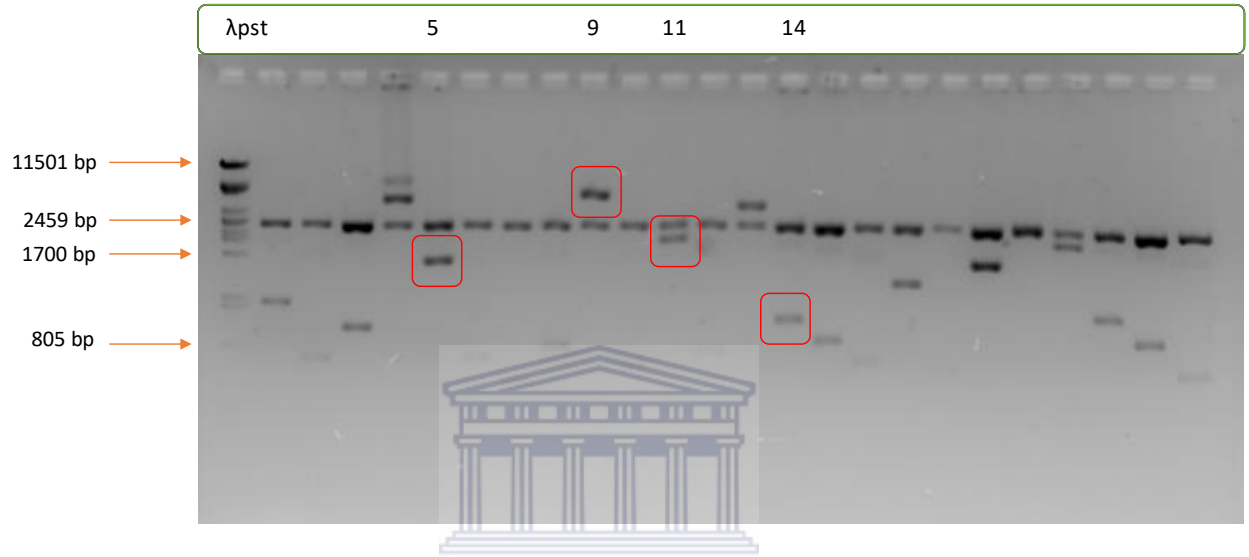

Figure 3.3: Agarose gel (1%) electrophoresis showing restriction patterns generated from digesting 24 recombinant pJET vector clones containing phage SMHB1 DNA fragments of random sizes. The clones were digested with *Swa*I. Lane one, Lambda phage DNA cut with *Pst*I used as a molecular marker. Sequenced clones are highlighted in red.

BLAST analysis of the sequenced inserts showed phage related proteins as summarized in Table 3.1 below. Clones 5 and 9 showed some similarity to hypothetical proteins on *Salinivibrio* sp. KP-1, which, although unknown at the time, contains a prophage very similar to phage SMHB1. Clone 11 was a fragment of an ORF encoding the terminase gene which has been used as a phylogenetic marker for establishing phage relatedness, as described in Chapter one.

Table 3.1: BLAST analysis of the 4 selected clones.

CLONE	BLASTn				BLASTx			
	TOP HIT	% QUERY COVER	E-VALUE	% IDENTITY	TOP HIT	% QUERY COVER	E--VALUE	% IDENTITY
5 (1.5 kb)	*Vibrio* phage PV94 complete genome	92	2e-110	72	Hypothetical protein [*Salinivibrio sp*. KP-1]	54	4e-110	100
9 (3kb)	*Enterobacter aerogenes* KCTC 2190, complete genome	12	4e-18	80	Hypothetical protein [*Salinivibrio sp*. KP-1]	36	5e-60	97
11 (2kb)	*Vibrio parahaemolyticus* BB22OP chromosome 1, complete sequence	90	2e-110	72	Terminase [*Photobacterium halotolerans*]	89	2e-119	72
14 (900bp)	*Vibrio vulnificus* strain 93U204 chromosome I, complete sequence	23	9e-17	77	Peptidase M15A [uncultured Mediterranean phage uvMED]	56	2e-41	64

Clone 14 contained sequence which was similar to a peptidase on the Mediterranean phage uvMED. Phage peptidases are functionally similar to lysin and lysozymes (O'Flaherty *et al*., 2005; Horgan *et al*., 2009). They are involved in compromising the cell wall peptidoglycan of the bacterial host to facilitate the exit of newly assembled phage particles (Becker *et al*., 2009).

Based on the low similarity of the clones describe above to hits on the database it was concluded that the phage was sufficiently different from known phages. The full phage genome was therefore sequenced.

3.3.5 Whole genome analysis

Sequence analysis report of the *Salinivibrio* species SS2 and SS3 is presented in Table 3.2. The raw reads were trimmed, de-multiplexed and the genomes assembled using CLC genomics version 6.5. For SS2, a total of 1,602, 658 paired end reads were generated with an average read length of 296.54. The final assembly also contained 151 contigs with an average length of 24,562 bp, the longest being 184, 246 bp. For SS3, a total of 1,900,628 paired end reads with an average length of 320.71 were generated. The final assembly contained 92 contigs with th longest being 398,091 bp and an average length of 37,723 bp. The final draft genomes were 3,708,917 bp (GC content 49.8%) and 3,470,512 bp (GC content 50.5%) for SS2 and SS3 respectively.

Basic annotation using the RAST server identified 3404 and 3169 coding sequences on the genomes of SS2 and SS3 respectively and 85 RNAs were also identified for SS2 compared to 81 RNAs on the genome of SS3. The bacteria SS3 was also found to contain a prophage of SMHB1 within its genome.

Table 3.2: Summary of the sequence data analysis for the whole genomes of Salinivibrio strains SS2 and SS3.

	Count		Average length		Total bases	
	SS2	**SS3**	**SS2**	**SS3**	**SS2**	**SS3**
Reads	2,094,188	2,469,124	253.86	261.36	531,629,716	645,319,050
Matched	1,870,748	2,214,316	248.98	257.51	465,783,426	570,202,406
Not matched	223,440	254,808	294.69	294.8	65,846,290	75,116,644
Contigs	151	92	24,562	37,722	3,708,917	3,470,512
Reads in pairs	1,602, 658	1,900,628	296.54	320.71		
Broken paired reads	267,213	312,898	283.97	269.75		

A summary of the *de novo* assembly of phage SMHB1 is provided in Table 3.3. The longest contig assembled for the whole genome of phage SMHB1 was 32845 bp in size, has a GC content of 50.8%, and contains 49 ORFs. No function could be assigned for 29 out of 49 genes identified which is typical of most phage genomes (Comeau *et al*., 2007; Hatfull, 2008).

Table 3.3: Summary of the sequence analysis report for the assembly of phage SMHB1

	Count	**Average length**	**Total bases**
Reads	399,221	289.21	115,460,503
Matched	364,766	289.54	105,613,275
Not matched	34,455	285.8	9,847,228
Contigs	481	3,675	1,767,914
Reads in pairs	250,662	387.84	
Broken paired reads	45,969	288.84	

A summary of the ORFs of phage SMHB1 is presented in (Table 3.4). The majority of annotated proteins are similar to phage K139, however, the whole genome seems to be interrupted by genes from other phages suggesting that the phage may have incorporated some genes during recombination events (Morris *et al*., 2008; Boyd *et al*., 2009).

Table 3.4: A summary of the BLASTx analysis of the ORFs of phage SMHB1. The + or – sign before the ORF number indicates on which strand the gene is found.

ORF #	ORF Position	Closest Phage BLAST Hit	Identities	Positives	E-value	Accession Number
	29916-29927	attL, ATAGGGAGACTG				
2/-	317-2503	Putative tail fiber protein (*Vibrio* phage K139)	171/273(63%)	210/273(77%)	6e-122	NP536667.1
3/-	2500-3114	Hypothetical protein (*Vibrio* phage K139)	83/198(42%)	112/198(56%)	3e-51	NP536666.1
4/-	3101- 4288	Hypothetical protein (*Vibrio* phage K139)	154/396(39%)	236/396(59%)	8e-104	NP536665.1
5/-	4281-4610	Hypothetical protein (*Vibrio* phage K139)	41/95(43%)	53/95(55%)	4e-21	NP536664.1
6/-	4607-6733	Putative tail length determinator (*Vibrio* phage K139)	245/482(51%)	327/482(67%)	9e-147	NP536663.1
7/-	6922-7188	Hypothetical protein (*Aeromonas* phage phiO18P)	33/84(39%)	51/84(60%)	6e-18	YP001285659.1
8/-	7185-7433	Hypothetical protein (*Shewanella halifaxensis*)	35/79(44%)	50/79(63%)	1e-21	WP012276442.1

9/-	7434-8018	Hypothetical protein (*Pseudomonas* phage vB_PaeM_PAO1_Ab27)	57/138(41%)	83/138(60%)	8e-21	YP009124372.1
10/-	8021-8434	Hypothetical protein (*Pseudoalteromonas* phage pYD6-A)	58/130(45%)	77/130(59%)	3e-31	YP007674297.1
11/-	8439-8645	Hypothetical; PP_00010			0.0	
12/-	8664-9119	Putative tail tube protein (*Vibrio* phage K139)	53/147(36%)	77/147(52%)	7e-31	NP536657.1
13/-	9119-10240	Putative tail sheath protein (*Vibrio* phage K139)	153/371(41%)	211/371(56%)	2e-86	NP536656.1
14/-	10243-10899	Putative tail completion protein (*Vibrio* phage K139)	65/197(33%)	110/197(55%)	4e-27	YP008766843.1
15/-	10883-11359	Hypothetical protein (*Vibrio* phage K139)	41/136(30%)	72/136(52%)	4e-16	NP536653.1
16/-	11359-11766	Putative head completion protein (*Vibrio* phage K139)	49/144(34%)	75/144(52%)	5e-20	NP536652.1
17/-	11865-12581	Terminase endonuclease subunit (*Vibrio* phage K139)	109/210(52%)	154/210(73%)	7e-83	NP536651.1
18/-	12584-13636	Capsid protein (*Vibrio* phage K139)	132/301(44%)	186/301(61%)	5e-89	NP536650.1
19/-	13636-14595	Capsid-scaffolding protein	142/318(45%)	191/318(60%)	4e-76	NP536649.1

		(*Vibrio* phage K139)				
20/+	14777-16570	Terminase ATPase subunit (*Vibrio* phage K139)	436/611(71%)	498/611(81%)	0.0	NP536648.1
21/+	16552-17610	Putative capsid portal protein (*Vibrio* phage K139)	215/333(65%)	258/333(77%)	7e-161	NP536647.1
22/+	17669-18106	Zinc-finger protein (*Vibrio* phage K139)	42/82(51%)	52/82(63%)	2e-20	NP536645.1
23/-	18088-18402	Hypothetical protein; PP_00022			0.0	
24/-	18399-18848	Hypothetical protein; PP_00023			0.0	
25/-	18866-19072	Hypothetical protein; PP_00024			0.0	
26/-	19065-19460	Hypothetical protein; PP_00025			0.0	
27/-	19616-19777	Hypothetical protein; PP_00026			0.0	
28/-	19901-20026	Hypothetical protein; PP_00027			0.0	
29/+	20007-20342	HTH-type transcriptional regulator (*Bacillus* phage vB_BhaS-171)	19/50(38%)	28/50(56%)	9e-04	ALY08093.1
30/-	20375-20719	Hypothetical protein; PP_00029			0.0	
31/-	20730-21284	Putative exonuclease (*Vibrio* phage vB_VpaM_MAR)	147/182(81%)	170/182(93%)	4e-115	YP007112509.1
32/-	21294-23726	Replication protein	363/767(47%)	496/767(64%)	0.0	NP536641.1

		(*Vibrio* phage K139)				
33/-	23723-23959	Hypothetical protein (*Vibrio* phage K139)	36/72(50%)	51/72(70%)	3e-22	NP536638.1
34/-	23968-24429	Hypothetical protein; PP_00033			0.0	
35/-	24434-24550	Hypothetical protein; PP_00034			0.0	
36/-	24560-24769	Hypothetical protein; PP_00035			0.0	
37/-	24834-25238	Hypothetical protein (*Salinivibrio* sp. KP-1)	120/134(90%)	126/134(94%)	6e-77	WP046074330.1
38/-	25248-25511	Hypothetical protein (*Enterobacteria* phage mEp460)	35/84(42%)	54/84(64%)	4e-23	YP007112099.1
39/-	25525-26046	CII protein (*Vibrio* phage K139)	58/173(34%)	96/173(55%)	5e-35	NP536633.1
40/-	26074-26232	Hypothetical protein; PP_00039			0.0	
41/-	26235-26435	Cox protein (*Pasteurella* phage F108)	21/53(40%)	34/53(64%)	5e-16	YP654715.1
42/+	26501-27127	CI protein (*Vibrio* phage K139)	73/211(35%)	116/211(54%)	4e-34	NP536631.1
43/+	27677-28186	Hypothetical protein (*Lactobacillus* phage Lv-1)	31/88(35%)	48/88(54%)	1e-05	YP002455813.1

44/-	28235-28723	Hypothetical protein; PP_00043			0.0	
45/-	28826-29890	Integrase (*Enterobacteria* phage mEp460)	196/356(55%)	253/356(71%)	3e-140	YP007112096.1
46/-	30377-30496	Hypothetical protein; PP_00045			0.0	
47/-	30579-30704	Hypothetical protein; PP_00046			0.0	
48/-	30731-31462	Hypothetical protein, ORF 35 (*Haemophilus* phage HP2)	94/299(31%)	144/299(48%)	3e-37	NP536842.1
49/-	31453-32139	Hypothetical protein (*Vibrio* phage K139)	69/219(32%)	123/219(56%)	5e-39	NP536671.1
50/-	32136-32681	Hypothetical protein (*Vibrio* phage K139)	60/156(38%)	88/156(56%)	2e-27	NP536670.1
	30113-30124	attR , ATAGGGAGACTG				

The most similar described phage to SMHB1 is the lysogenic, *Vibrio vulnificus* –infecting, phage PV94 (Pryshliak *et al*., 2014), with which it shares 69% identity over 36% of its genome at a nucleotide level. However, the SMHB1 sequence is most similar to a prophage on the recently sequenced bacterium, *Salinivibrio* sp KP-1, sharing an identity of 94% over 56% of the entire genome.

Local translated nucleotide alignments (tBLASTx) of the genomes of these phages are represented below (Figure 3.4A-C). Even though it shares the highest nucleotide identity with the prophage on the KP-1 genome and high identity with PV94, phage SMHB1 is more closely related to *Vibrio* phage K139, a temperate myovirus of *Vibrio cholerae*, at the protein level (Figure 3.4D).

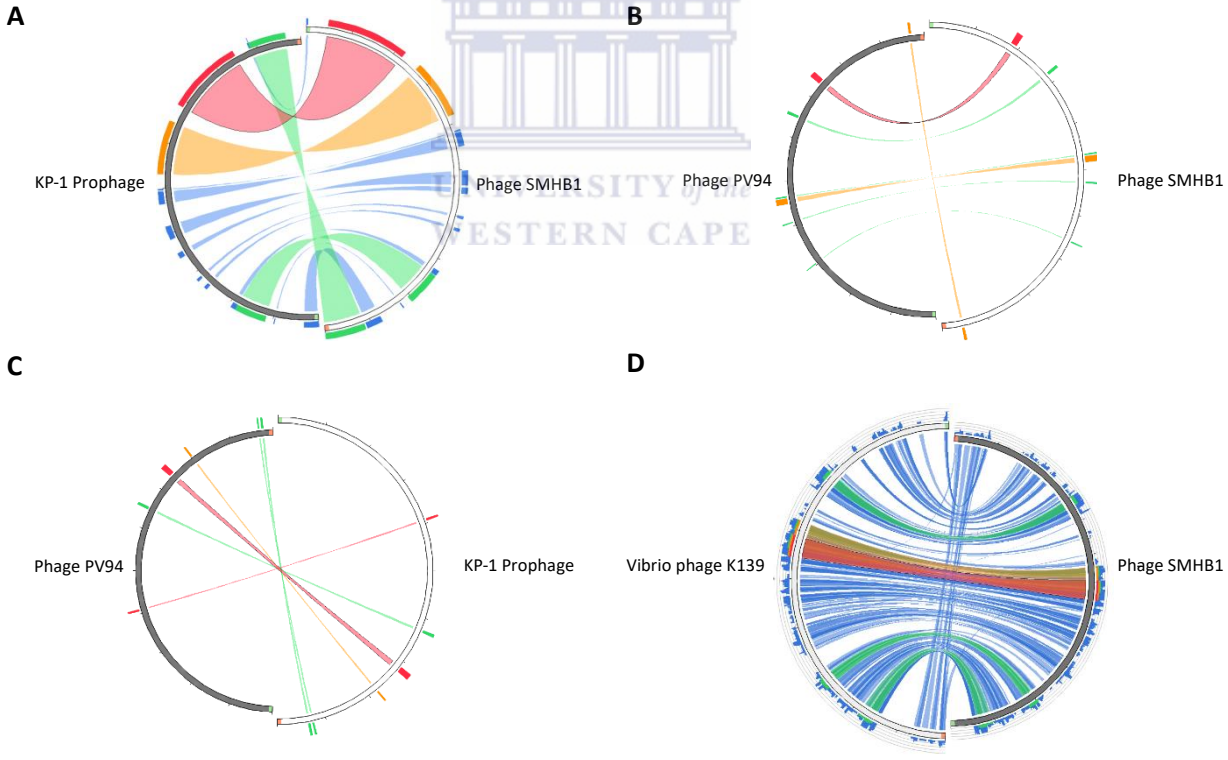

Figure 3.4: Ideograms representing local alignments of BLAST results of compared genomes. Ribbons represent regions which are similar. The degree of similarity is represented by a four- colour scheme: blue (<25%), green (25-49%), orange (50-74%) and red (>75%). Inverted ribbons represent sequences that align on the opposite strand (Darzentas, 2010). A-C. Comparison at a nucleotide level. D. Comparison at a protein level.

Interestingly, none of these three phages shared any significant nucleotide or protein similarity to the podovirus CW02, one of the two published *Salinivibrio* infecting phages (Shen *et al*., 2012). The second *Salinivibrio* phage is the myovirus UTAK (Goel *et al*., 1996) but the genome is unavailable and the phage is currently propagated on *Vibrio hispanica* in culture collections. Therefore, the host has not been identified with certainty. This could indicate that phage SMHB1 may infect *Vibrio* species.

Tetranucleotide usage deviation (TUD) patterns have been shown to be useful in determining ancestral relationships between phages, hosts and other closely related phages (Pride *et al*., 2006). Such relationships are represented as a Pearson's correlation coefficient, which is representative of the strength of the relationship between the two variables (phages in this case). Values usually range between -1 and +1, where a value closer to +1 indicates a high possibility of a positive relationship while -1 to 0 indicate negative or no relationship. Therefore, assuming that the relationship between phage SMHB1 and the bacterial host, SS3 gives a value of 1, the analysis of SMHB1 with respect to the KP-1 prophage gave a coefficient of 0.89649 compared to 0.38668 for phage CW02 (Table 3.5). This suggests that SMHB1 and KP-1 prophage are more likely to infect the same bacterial host compared to CW02. The *Vibrio* phages PV94 and VPUSM8 gave coefficients of 0.69811 and 0.65004 respectively, when compared to phage SMHB1, showing that they are less likely to or have previously infected *Salinivibrio* species.

Table 3.5: Summary of the Pearson's correlation coefficients showing the relationship between SMHB1 and similar phages.

	SMHB1	KP-1	CW02	PV94	VPUSM8
SMHB1	1	0.89649	0.38668	0.69811	0.65004

The nucleotide usage pattern of SMHB1 was compared to the three *Salinivibrio* genomes available on the NCBI database yielding the coefficients 0.74192, 0.74504, 0.77010, 0.67918 for *Salinivibrio socompensis*. S35 (NZAQOE00000000.1), *Salinivibrio costicola* ATCC 33508 (NZAQOF00000000.1), *Salinivibrio* KP-1 (NZLAQR00000000.1) (Table 3.6). *Vibrio vulnificus* YJ016 (GCA000009745.1) was included due to the nucleotide similarity of its infective phage PV94 to SMHB1.

Table 3.6: Summary of the Pearson's correlation coefficients showing the relationship between SMHB1 and Salinivibrio bacterial species.

	Salinivibrio socompensis S35	*Salinivibrio costicola* ATCC 33508	*Salinivibrio* KP-1	*Vibrio vulnificus* YJ016
SMHB1	0.74192	0.74504	0.77010	0.67918

Based on the assumption that TUD analysis suggests a level of phages adapting their genomes to that of the host (Pride *et al*., 2006), the relatively high values derived here suggest that SMHB1 may have recently infected *Salinivibrio* KP-1 or any of these other species. It is also possible that a *Vibrio* or another host exists in nature.

It has been shown that the replication mechanism directly causes asymmetries in nucleotide base composition referred to as skews (Lobry, 1996). The base composition of each gene is dependent on its position on the leading or lagging replication strand, and a statistical analysis of the cumulative effect on the entire genome shows definite breakpoints which correspond to the origin and terminus of replication (Tillier & Collins, 2000; Necşulea & Lobry, 2007). A GC skew analysis of the phage genome (Figure 3.5) suggested that the origin of replication lies at position bp 23101 which is within ORF 32, predicted to encode a putative replication protein. The replication terminus is also suggested to lie at 14401 bp within ORF 19 encoding the

capsid-scaffolding protein which has been shown in the *Enterobacteria* phage P2 of *E. coli* to cleave itself and become the scaffold protein on which the head of the bacteriophage is built (Marchler-Bauer *et al*., 2015). This suggests that phage SMHB1 may assemble new viral particles in a similar way to phage P2.

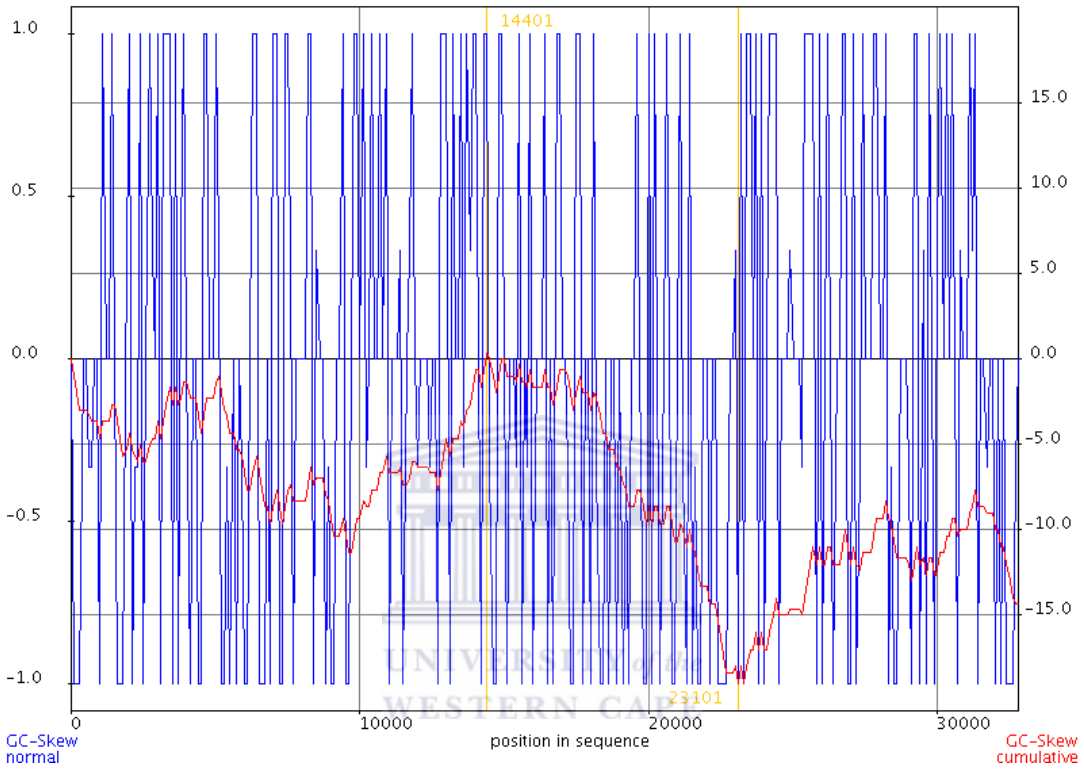

Figure 3.5: A GC skew analysis of the genome of phage SMHB1 as a tool to predict the putative replication origin (ori) and termination sites (ter) calculated using a window size of 1000 bp and a step size of 100 bp.

Two repeat regions greater than 10 base pairs were found in phage SMHB1, a 19 bp fragment, CGCGGGTCGCTTGTCCAGT, located at either ends of the linear genome (bp 1- 19 and bp 32827- 32845). No similarity to known sequences was found using BLAST analysis. The sequence probably represents the physical end of the phage genome which is repeated as a consequence of the assembly algorithm. Where the DNA is concatenated or there are several overlapping contigs, the assembler may have recognized the region as repeat sequence when it is in fact the 5' beginning of the next genome. Scanning the whole genome of SS3 for this

repeat sequence showed that when the phage is integrated, it is found only once within the phage sequence. This supports the argument that the repetition of the sequence was a consequence of the assembly algorithm.

The other interesting repeat region is the 12 bp sequence ATAGGGAGACTG (bp 29916-29927, 30113-30124). The first region lies 26 bps downstream of the integrase gene (ORF 45) just before the tRNA dihydrouridine synthase (DusA), a host bacterial gene involved in the post translational reduction of uridine (Kasprzak *et al*., 2012; Farrugia *et al*., 2015). Analysis of the phage when integrated within the genome of SS3 showed that the second repeat sequence lies within a fragment downstream of the lipopolysaccharide A (LPS A) gene of SS3 and upstream a suite of phage hypothetical proteins (**Figure 3.6**). BLASTx analysis of this fragment showed that it is still part of the DusA gene, most likely the N- terminus.

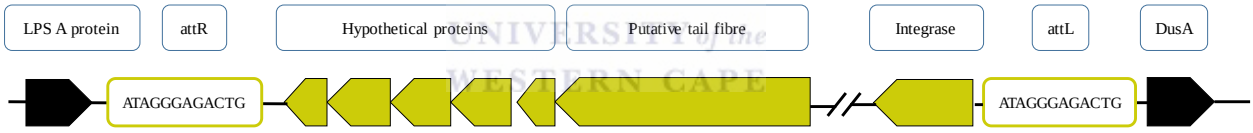

Figure 3.6: Graphical representation of the integration site of phage SMHB1 within the genome of the host, SS3. Phage genes are coloured yellow while black arrows represent bacterial genes.

This suggested that the repeat sequence could represent the hybrid attachment sites, attR and attL, together forming a 24 bp phage attachment (attP) site used for integrating into the genome of the host. This also confirmed that the phage integrates within the tRNA dihydrouridine synthase gene of the host, a phenomenon that has been observed for prophages infecting over 200 bacteria (Farrugia *et al*., 2015). The study also observed that the integrase genes of these prophages belong to a novel class that encode functions such as metabolic augmentation and heavy metal resistance (Farrugia *et al*., 2015). This suggests that phage SMHB1 might play a

role in the survival of *Salinivibrio* species in heavy metal rich environments (Gorriti *et al*., 2014) and would form the basis of an interesting future study. The repeat region was found twice within the genome of KP-1 prophage, suggesting both SMHB1 and the KP-1 prophage might infect the same host.

Within the mature phage, the 12 bp repeat region is separated by a 185 bps nucleotide sequence which BLASTx analysis showed to encode a fragment of the same tRNA dihydrouridine synthase (DusA) identified on the SS3 genome, (Figure 3.7). This suggests that the phage incorporates a fragment of the host genome during excision.

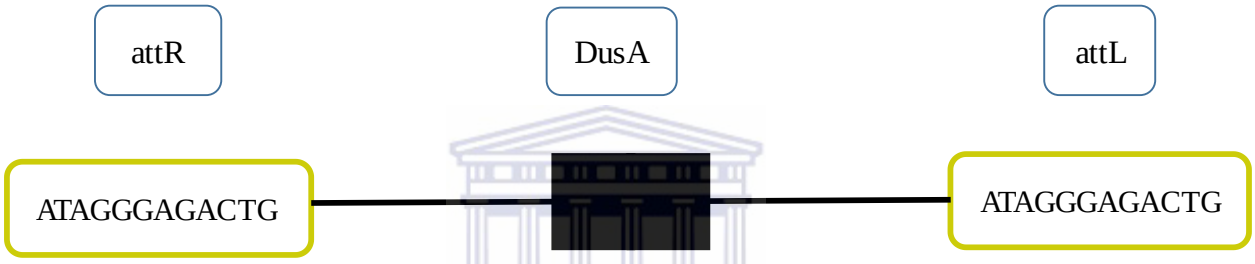

Figure 3.7: Arrangement of the phage integration site with respect to the fragment of the tRNA dihydrouridine synthase (DusA) gene excised during conversion from the lysogenic to the lytic cycle.

The ability of phages to excise fragments of the host DNA and package it into heads of newly assembled phages is referred to as transduction. Lambda phage has been shown to transduce galE which may confer the ability to metabolize galactose on a susceptible host (Sato & Campbell, 1970) and the recA gene which is essential for DNA repair and maintenance (McEntee, 1976). Given that this is the first report of phage SMHB1incorporating such foreign DNA and that it seems to incorporate a fragment of rather than the entire DusA gene, it cannot be ascertained whether it is capable of generalized (incorporation of random genes) transduction, specialized (incorporation of specific genes) transduction, or both as observed in the *Salmonella typhimurium* infecting phage P22 (Kwoh & Kemper, 1978). Also, the function of dihydrouridine in tRNA is not completely understood but it has been suggested that it

promotes structural flexibility by destabilizing the C3′-endoribose conformation associated with base-stacked RNA (Dalluge *et al*., 1996; Yu *et al*., 2011). The phage might require the gene for efficient transcription and translation during the assembly of new viruses using the host's machinery. Transduction ability has been explored in biotechnology for the genetic manipulation (Bardarov *et al*., 2002) and development of vaccines against *Mycobacterium tuberculosis* (Tufariello *et al*., 2014). It would be worth investigating if phage SMHB1 can be similarly engineered.

3.3.5.1 Lysis/Lysogeny

No lysis associated genes were readily identifiable on phage SMHB1 but a lysogeny module was found. The arrangement of genes within the lysogeny module of most phages is thought to follow a similar pattern- attL-Integrase-CI-Cro-Excisionase (Brüssow & Desiere, 2001; Romero *et al*., 2004), however, phage SMHB1 does not completely follow that pattern, with the genes arranged- attL- Integrase- Hyp-Hyp-CI-Cox- Hyp- CII (Figure 3.8).

The integrase protein (ORF 45) is required for both integration and excision of the prophage (Gottesman & Yarmolinsky, 1968). During integration, it splices together the phage (*attP*) and bacterial (*attB*) attachment sites by binding to specific sites on the phage DNA, folding it into a complex structure ready for subsequent recombination (Echols & Court, 1971; Weisberg & Landy, 1983). The same folding action causes the detachment of the *attP* site from the chromosome at the onset of the lytic cycle (Weisberg & Landy, 1983). The presence of this gene made phage SMHB1 and its host the ideal system for optimizing the hybridization experiments described in the next Chapter. The host is naturally occurring in the saline site as a lysogen. This therefore provided a system where the optimized FISH methodology could be applied a known phage-host pair from an extreme environment (Chapter four).

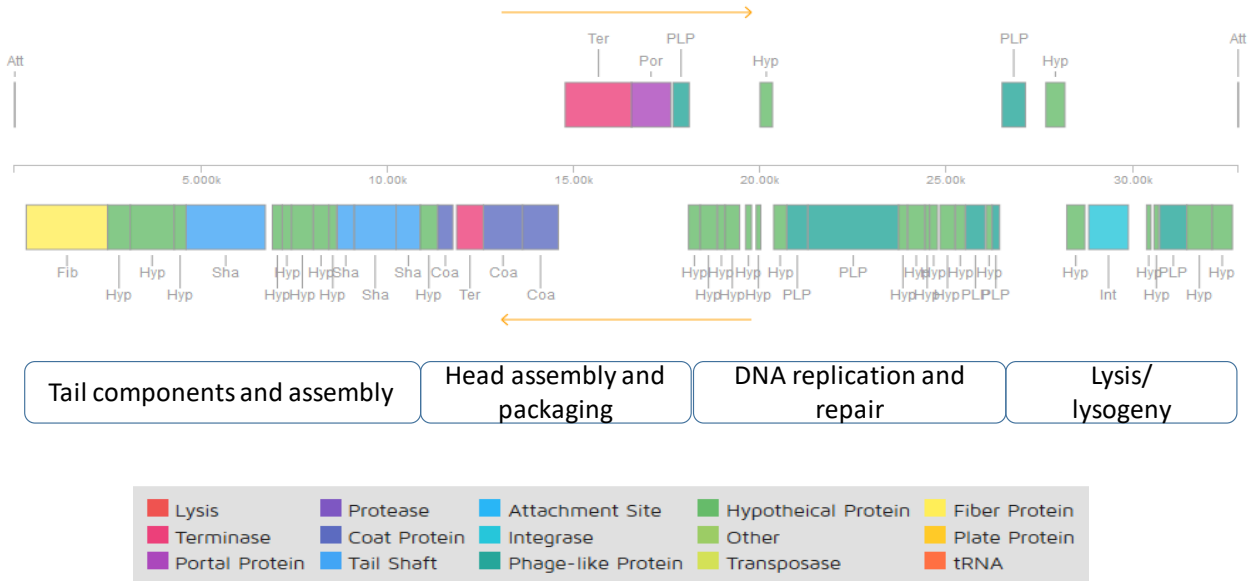

Figure 3.8: An overview of the genome of phage SMHB1 using the PHASTER tool (Arndt *et al*., 2016). The arrows are indicative of the orientation of the ORFs.

Both the CI repressor (ORF 42) and CII proteins (ORF 39) are known to work in conjunction with other proteins which are not readily identifiable in phage SMBH1. The repressor protein works opposite to the *cro* protein for control of an operator region which is the genetic switch that "decides" between lysogeny or lysis (Figure 3.9) (Aggarwal *et al*., 1988). If the competition, a complicated cascade which is not yet fully understood, is won by the repressor protein, the transcription mechanism of the *cro* gene is blocked, late phage genes are not expressed and lysogeny occurs (Campanizzi, 2014).

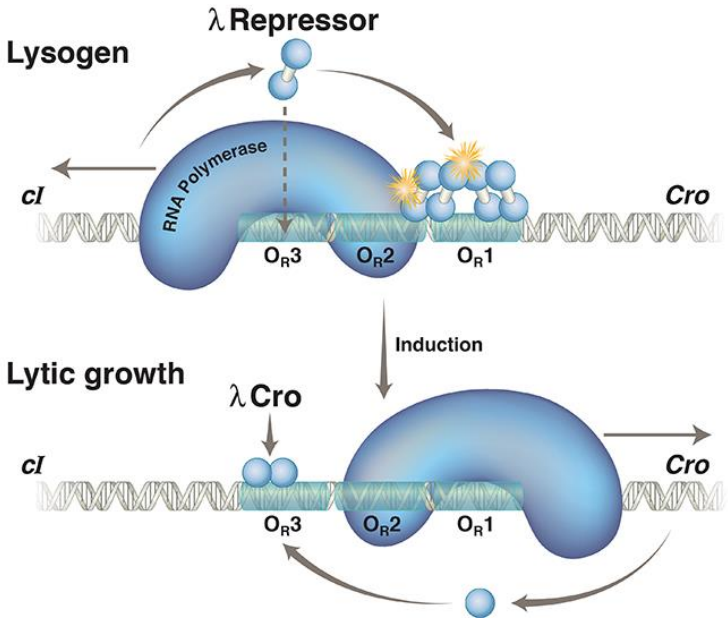

Figure 3.9: Graphical representation of the roles of the repressor cro genes in lysis/lysogeny switch. The switch is proposed to involve a complex network of interactions which include a double negative circuit involving the repressor and the protein *cro*, interactions between proteins separated by 3,000 base pairs which result in the formation of a large DNA loop, and other positive and negative feedback not completely understood (Campanizzi, 2014).

Lysis therefore only occurs when the *cro* gene blocks the CI repressor gene from being transcribed (Aggarwal *et al*., 1988; Mondragon & Harrison, 1991). The CII protein functions in concert with CIII to activate promoters Pi and Pe which are responsible for the transcription of the integrase and repressor genes respectively (Shotland *et al*., 1997; Parua *et al*., 2010). The inability to annotate the *cro* and CIII proteins for phage SMHB1 might be a consequence of the relatively small size of currently available phage databases or a regulatory network unlike that found in phage lambda. It is also possible that the functions of both genes are performed by one or more of the hypothetical proteins within the lysogeny module.

The Cox protein (ORF 41) was characterized in *E. coli* phage P2 and shown to be involved in three major processes, namely the excision of the integrated phage genome (Lindahl & Sunshine, 1972), transcriptional activation of the late PII promoter for satellite phage P4 (Saha *et al*., 1989) and the repression of the P2 Pc promoter that controls the expression of the

integrase and repressor genes (Saha *et al*., 1987). Interestingly, the Cox protein in SMHB1 is most similar to that found in the *Pasteurella* phage F108 (79% query cover at 40% identity), also a myovirus and one of the two known non-*E. coli* infecting phages capable of transduction (Campoy *et al*., 2006). This further supports the possibility of SMHB1 also being capable of packaging bacterial DNA into the viral envelope.

It is also interesting to note that the Cox protein, which functions similarly to an excisionase (Lindahl & Sunshine, 1972) is unconventionally located between the transcriptional repressor, CI and the transcriptional activator, CII . It would also be interesting to study the significance, if any, of the unconventional arrangement of genes within the lysogeny module of SMHB1.

3.3.5.2 DNA replication, packaging and repair

Repeat regions less than 10 base pairs are known to be associated with regions involved with the initiation of DNA replication, or sites involved in gene regulation and the termination of transcription (Østergaard *et al*., 2001; Blatny *et al*., 2004; Pride *et al*., 2006). Five such regions were identified within the genome of phage SMHB1: CTTTGTCG (bp 41-48, 248-255, 22043-22050), TCACTAAT (bp 139- 146, 160 -167, 21078- 21085), TATTTCTT (bp 243-250, 267-274, 17552 -17559, 24137 -24144, 28827 -28834), TTGTCGTGT (bp 250 -258, 289- 297, 14804 -14812), CTTCTTGTTC (bp 131- 140, 212-221). CTTTGTCG is found once within the replication gene (ORF 32) and TTGTCGTGT lies within the terminase gene (ORF 17) which is involved in DNA packaging (Feiss & Catalano, 2005; Rao & Feiss, 2015). These repeat regions might be involved in replication while the rest are associated with other processes.

The replication protein (ORF 32) is involved in the bi- directional replication of DNA (Sanger *et al*., 1982) while the exonuclease (ORF 31) is thought to catalyze a number of reactions that include the formation of heteroduplex regions, exposure of complementary sequences and the

elimination of redundant branches (Cassuto *et al*., 1971). The terminase endonuclease subunit (ORF 17) and terminase ATPase subunit (ORF 20) encode genes responsible for initiation of packaging viral DNA by cutting the DNA concatemer. The packaging process occurs by DNA translocation and in some phages, e.g. lambda phage, the terminase gene has been shown to actively support the process of coupling ATP hydrolysis to DNA translocation (Black, 1995).

The terminase subunit has been demonstrated as not being only useful as a phylogenetic marker (Liu *et al*., 2011; Shan *et al*., 2012), but also as an indicator of the type of genetic material packaging mechanism employed by the phage (Black, 1995; Feiss & Catalano, 2005). Generally, the mechanisms of DNA replication in phages result in the build-up of DNA concatemers which cause mature phage DNAs to be joined together in a head-to-head configuration through terminal repetitions (Black, 1995; Feiss & Catalano, 2005). Each concatemer eventually becomes the substrate for DNA encapsulation into previously assembled heads. This mechanism is specific for each phage but is facilitated by the terminase gene. Phages with similar terminase genes have been shown to package DNA in the same manner (Hsiao & Black, 1978; Black, 1995; Feiss & Catalano, 2005). Terminase phylogenetic analysis was therefore similarly employed here to measure the degree of relatedness of the closest phage hits on the NCBI database which might suggest the DNA packaging mechanism phage SMHB1 employs. A BLAST analysis of the large terminase subunit of SMHB1 was conducted followed by an analysis of the top phage terminase hits (Figure 3.10). Other characterized phages with described packaging mechanisms were included for context. The accession numbers for the top phage terminase hits are listed in Appendix A.

The phylogenetic analysis showed that SMHB1 clusters with phages F108, HP1, HP2, PV94, VD1 K139 and VPSUM8 and the KP-1 prophage. Of this group, the DNA packaging mechanisms of HP1, K139 and PV94 have been characterized and determined to be *Cos*-packaging (Linderoth *et al*., 1991; Reidl & Mekalanos, 1995; Pryshliak *et al*., 2014).

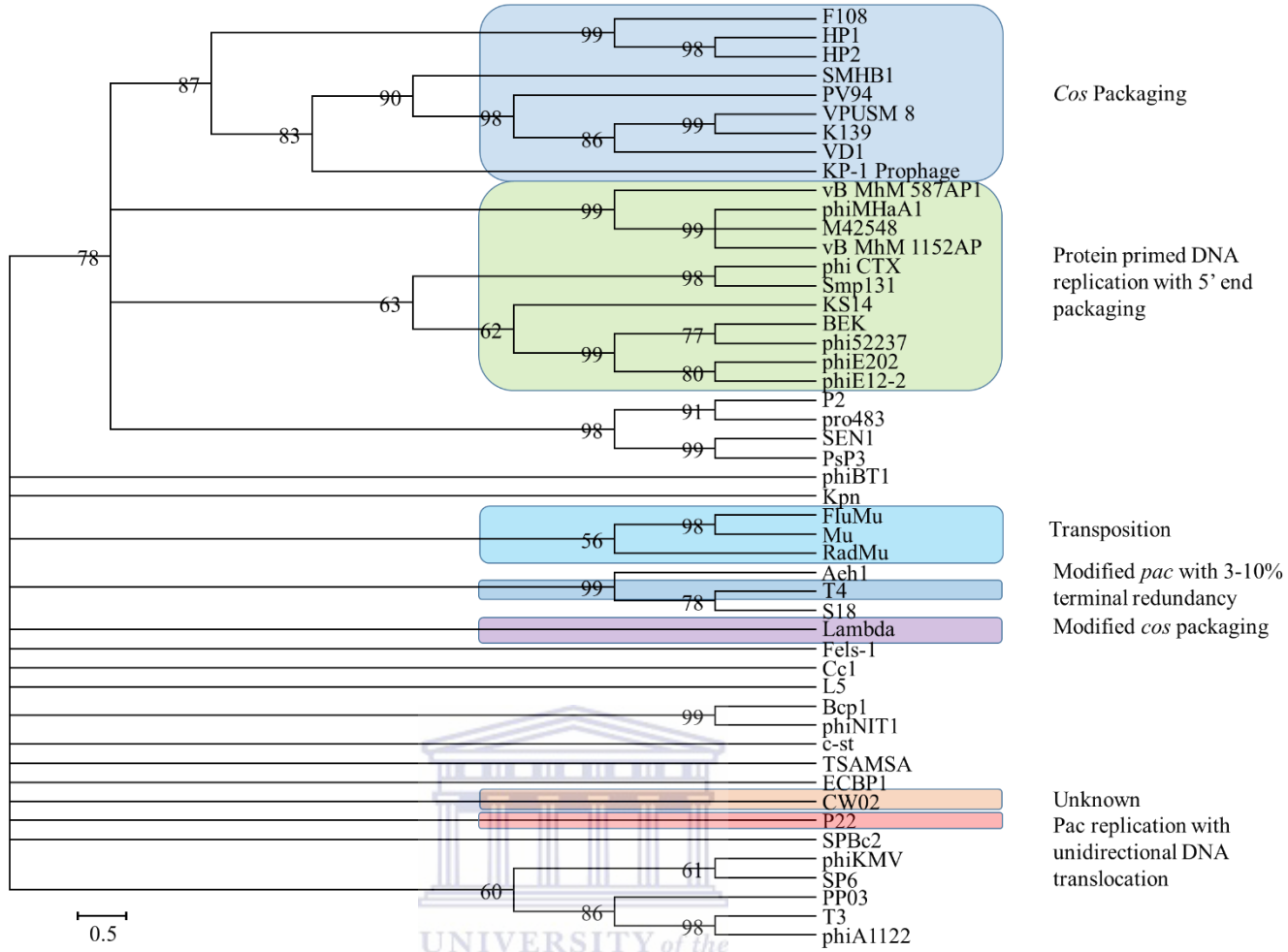

Figure 3.10: Phylogenetic tree comparing phage terminases. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Jones *et al*., 1992). The tree with the highest log likelihood (-12544.0611) is shown. The percentage of trees in which the associated taxa clustered together is shown above the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The analysis involved 49 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 160 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura *et al*., 2013).

The terminase gene of phage HP1 has been shown to catalyze the staggered cleavages that produce the cohesive ends (*cos* sites) of the mature linear DNA (Linderoth *et al*., 1991; Esposito *et al*., 1996). Phage K139 has been shown to also possess a *cos* site similar to P2-like phages but from which terminal redundant ends diverge, suggesting that after infection, the phage DNA might circularize by either using the *cos* site or a recombination of these terminally

redundant ends (Reidl & Mekalanos, 1995; Kapfhammer *et al*., 2002). Phage PV94 was shown to achieve this circularization using a *cos* site (Pryshliak *et al*., 2014). Given the degree of relatedness to the terminase of phage PV94, it is possible that the DNA packaging mechanism of phage SMHB1 is similar. The possession of a *cos* dependent mode of replication also suggests that SMHB1 could be engineered as a cloning vector for the preparation of fosmid libraries.

Introns have been found in phages to be involved in DNA manipulation, for example, T4 phage has three self-splicing introns in genes in *de novo* deoxyribonucleotide biosynthesis (Sandegren & Sjöberg, 2007) and phage Twort is thought to use introns as a means of encoding multiple RNAs within one gene (Landthaler & Shub, 1999). No intron was found on the genome of phage SMHB1 but it encodes a Zinc finger like protein (ORF 22). Zinc finger (ZnF) like proteins are small but functional domains that are coordinated by at least one Zinc ion (Laity *et al*., 2001). They are structurally diverse and therefore able to function in processes such as DNA recognition, protein folding and assembly, lipid binding, and transcriptional activation (McNamara, 2000; Laity *et al*., 2001). A homolog of ORF 22 is also found in *Vibrio cholerae* phage K139 (59% identity at 97% query cover, E-value 2e-19) where it was suggested to be involved in the regulation of the expression of late genes (Kapfhammer *et al*., 2002). In phage SMHB1, the Zinc finger like protein might have a similar role, perhaps working in tandem with ORF 29 which is similar to the HTH-type transcriptional regulator found in *Bacillus* phage vB_BhaS-171. Alternatively, the Zinc finger gene might be used by the phage to manipulate the integrase gene recognition sequence as demonstrated for HIV-1 (Sakkhachornphop *et al*., 2009). The specificity of Zinc finger like proteins has been explored in biotechnology by engineering specific ZnF motifs in tandem within a protein to recognize specific DNA sequences, while another domain on the protein performs a desired function (Urnov *et al*., 2010).

3.3.5.3 Virion morphogenesis

The protein coat of phage SMHB1 is encoded by at least four genes: the putative head completion protein (ORF 16) which allows for complete head assembly by making the newly packaged DNA resistant to bacterial DNAse (Linderoth *et al*., 1991); the capsid protein (ORF 18) and the capsid-scaffolding protein (ORF 19) which binds the capsid protein, forming a cylindrical portal through which the viral genetic material is packaged into the phage head during assembly; and the putative capsid portal protein (ORF 21). These genes work in tandem during viral assembly and are involved in maintaining capsid pressure for efficient ejection of the genetic material during host infection (Murialdo, 1977; Katsura & Kobayashi, 1990; Dokland & Murialdo, 1993).

The genomic organization of head and tail assembly modules of phage SMHB1 shows a very close relationship to phage K139 with 11 characterized genes showing significant homology (Table 3.4). A BLASTp analysis of the capsid protein of SMHB1 (ORF 18) revealed a domain found in most known P2-like phages including phages K139, HP1and HP2, and which has been shown to be highly conserved in terms of similarity and structure (Kapfhammer *et al*. 2002). This is thought to be a reflection of the capsid of P2-like phages as just a DNA container which requires little or no adaptation to different hosts (Kapfhammer *et al*. 2002). Given the homology to P2-like phages, the capsid of phage SMHB1 most likely plays no role in determining susceptible hosts. The tail component and assembly module contains 13 genes, 8 of which could not be assigned any function. The five that could be annotated include the putative tail fiber protein (ORF 2), putative tail length determinator (ORF 6), putative tail tube protein (ORF 12), putative tail sheath protein (ORF 13), and putative tail completion protein (ORF 14). Although ORFs 3 and 4 were identified as hypothetical proteins, they contain the *E. coli* phage P2 proteins I and baseplate protein J domains respectively. In phage P2, the protein J lies at the edge of the baseplate and is thought to be horizontally transferred (Haggard-

Ljungquist *et al*., 1992; Haggard-Ljungquist *et al*., 1995). It has also been suggested that the presence of these domains across phage families shows that recombination events occur at unprecedented levels and that the host range of these phages is altered under selective pressure (Haggard-Ljungquist *et al*., 1992). It might therefore be interesting to test if phages P2 and SMHB1 have a similar host range. Repeat regions are found in ORFs 2, 6 and 13 are thought to play a significant role in determining tail length (Belcaid *et al*., 2011).

3.3.5.4 Bacterial genes

Phages are known to carry bacterial genes that may be beneficial to efficient propagation in the host (Canchaya *et al*., 2004) and phage SMHB1 appears to have a protein from *Shewanella halifaxensis* (ORF 8). However, the function in the bacteria is currently unknown.

3.3.5.5 Possible mechanism of phage SMHB1 resistance in SS2

Apart from novelty, the genomes of both *Salinivibrio* species isolated from the saline site (Chapter two) were sequenced in an attempt to identify possible genetic differences that may explain why SS3 was susceptible to phage SMHB1 but SS2 was not. One of such possible defence mechanisms is the presence of CRISPR regions (Sorek *et al*., 2008; Bondy-Denomy *et al*., 2013). In microbial hosts, CRISPR loci usually contain a combination of CRISPR-associated (Cas) genes and non-coding RNA elements which are responsible for the target specificity of CRISPR-mediated nucleic acid cleavage (Bikard & Marraffini, 2013; Rath *et al*., 2015). A CRISPR locus will also have an array of direct repeats interspaced by short stretches of non-repetitive sequences referred to as spacers (Bikard & Marraffini, 2013). Spacers play an important role in target recognition and usually encode a fragment of a phage or genetic element the bacteria is resistant to (Hélène Deveau *et al*., 2008; Stern *et al*., 2012). Seven

CRISPR regions were identified in SS2 and 3 regions in SS3. The regions and the number of spacers associated with each are summarized in Table 3.7.

Table 3.7: CRISPRs and number of spacers identified in the genomes of *Salinivibrio* species SS2 and SS3 using CRISPRFinder (Grissa *et al*., 2007).

SS2		**SS3**	
CRISPR Region	Number of Spacers	CRISPR Region	Number of Spacers
GCTTCACGCTTAGCTTTCTCTTC (23 bp)	3	TTTCTAAGCTGCCTGTGCGGCA GTGAAC (28 bp)	36
TTTCTAAGCTGCCTGTGCGGCA GTGAAC (28 bp)	9	TTTCTAAGCTGCCTGTGCGGCA GTGAAC (28 bp)	41
TTTCTAAGCTGCCTGTGCGGCA GTGAAC (28 bp)	24	CGGTTCATCCCCGTGGGCACG GGGAACAC (29 bp)	48
TTTCTAAGCTGCCTATGCGGCA GTGAAC (28bp)	35		
TTTCTAAGCTGCCTATGCGGCA GTGAACA (29 bp)	3		
GTGTTCCCCGTGCCCACGGGGA TGAACCG (29 bp)	42		
GTCGCGCCTCCCGCAGGCGCGT GGATTGAAAC (32bp)	155		

A BLAST analysis of all the spacers identified in both bacteria against phage SMHB1 genome showed that seven were perfect matches to the genome. The seven spacers were all found within the genome of SS2. As previously described in *E. coli*, spacers are non-randomly

selected (Swarts *et al*., 2012) and the genes targeted by these SS2 spacers suggests specificity to important processes in the life cycle of the phage (Table 3.8).

Table 3.8: SS2 spacers with perfect matches to the genome of phage SMHB1

Spacer	Phage region	SMHB1 gene recognized
TTAGCTTCTGCGGGGTGGTTCGCGCAGCGGTAGC	8226-8259	Hypothetical protein (*Pseudoalteromonas* phage pYD6-A)
ACATCAACGGCGTGCCCTATCTGCGTGAAGACGAG	8685-8719	Putative tail tube protein
ATTGTTGCCTGCTGATCAGGTCAAAGGGCAAGTTA	13435-13469	Capsid protein
CGGTGTATATGCCATATGCGGCAATAGTACGCC	14768-14800	Terminase ATPase subunit
GACAAAAAATCGGTGTCGGCAACGACACCATC	21426-21457	Replication protein
AAAACCAAAACACGCGAATACGTGGGGCCGCT	22309-22340	Replication protein
TCGAATGGGTGGTGAACTACAACGCCGCATTC	23387-23418	Replication protein

These spacers specifically target genes involved phage DNA delivery into a susceptible host (tail tube protein) (Langlois *et al*., 2015), head formation (capsid protein) (Anastasia & Rossmann, 2011), the terminase gene which is responsible for DNA packaging into newly assembled phage heads (Feiss & Catalano, 2005) and three spacers specifically target the replication protein which is involved in synthesizing new phage genetic material (Østergaard

et al., 2001). Some of these genes, for example, the terminase, are conserved in similar phages (Rao & Feiss, 2015) which suggests that those phages may be susceptible to the CRISPR defence mechanism of SS2.

The presence of spacers recognizing phage SMHB1 suggests that the CRISPR system is at least one of the defence mechanisms through which SS2 evades infection by the phage. This also shows that it is likely that the phage successfully attaches to SS2 and injects its DNA which gets degraded, a case of abortive infection (Section 1.6.4).

3.4 CONCLUSION

In summary, phage SMHB1 was characterized as a myovirus with dimensions within the reported range. The whole genome of phage SMHB1 was sequenced. Regions of the genome appear conserved but the similarity of some of the genes to phages infecting bacteria from different order suggests its involvement in widespread recombination events. Despite this, the host range, although not exhaustively tested, appears to be limited to *Salinivibrio*. Putative functions assigned to 20 of the identified ORFs and the phage appears to display the classical modular arrangement while encoding unique genes such as the Zinc finger which may enhance the efficiency of DNA manipulation. There is some evidence to suggest that SMHB1 replicates using a *cos* site which would make it a candidate for engineering for packaging relatively large DNA fragments. The phage also has some interesting features such as the non-classical arrangement of genes within its lysogeny module, the possibility of being capable of transduction which could suggest its use as a DNA vector, the presence of a novel integrase which may confer heavy metal resistance capability on the host bacteria and the identification of a CRISPR defence mechanism in a non-susceptible host. With further investigation, biotechnological applications may be found for the novel phage.

4

FLUORESCENT LABELLING

4.1 INTRODUCTION

Fluorescent labelling offers a different approach to isolating novel phages compared to the traditional isolation technique mentioned in Chapter one. While culturing provides a direct link between the phage and its host, its limitations in terms of low throughput and difficulty in culturing most hosts raises the need for new methods. A number of such methods have been developed. For example, microfluidic digital PCR has been used to identify viral gene products and co-localized bacterial hosts but it is dependent on the availability of genetic information for both the virus and host for primer design (Tadmor *et al*., 2011). Immunomagnetic beads coated with fluorescently labelled viruses have also been used to detect *E. coli* using flow cytometry (Goodridge *et al*., 1999a; Goodridge *et al*., 1999b), however, the method is limited by its strain specificity and the need for high cell concentrations (Goodridge *et al*., 1999b). Also, the combination of fluorescent labelling of viruses with flow cytometry has been explored but the experiments were plagued by labelling of non-host cells (Ohno *et al*., 2012).

In this Chapter, we sought to improve on fluorescent labelling and in tandem with flow cytometry, attempt to identify and isolate novel phage-host pairs from environmental samples using culture-independent approaches. Two approaches to fluorescent labelling were explored namely viral tagging and fluorescence *in situ* hybridization (FISH). Both methods rely on fluorescence activated cell sorting (FACS), a specialized form of flow cytometry. These concepts are described below:

4.1.1 Fluorescence activated cell sorting (FACS)

FACS is capable of differentiating between cells within a heterogeneous mixture with almost 100% accuracy on the basis of fluorescence, light scattering ability and physical properties (Brussaard *et al*., 2000). The equipment has a nozzle with a small diameter such that cells of

interest are forced through one at a time and contained within a stream of sheath fluid (Figure 4.1). The nozzle is vibrated at a pre-determined optimal frequency and an attached laser scans the cells as they pass through. The cells scatter some of the laser light, and this parameter is used to both count the number of cells and measure their sizes. Where fluorescently labelled cells are used, light from the specific laser excites the dye and the emitted light is detected by a light detector or photomultiplier which relays the information to a computer. Each cell within the stream breaks off into single droplets after interrogation (Herzenberg *et al*., 2002; Herzenberg *et al*., 2002).

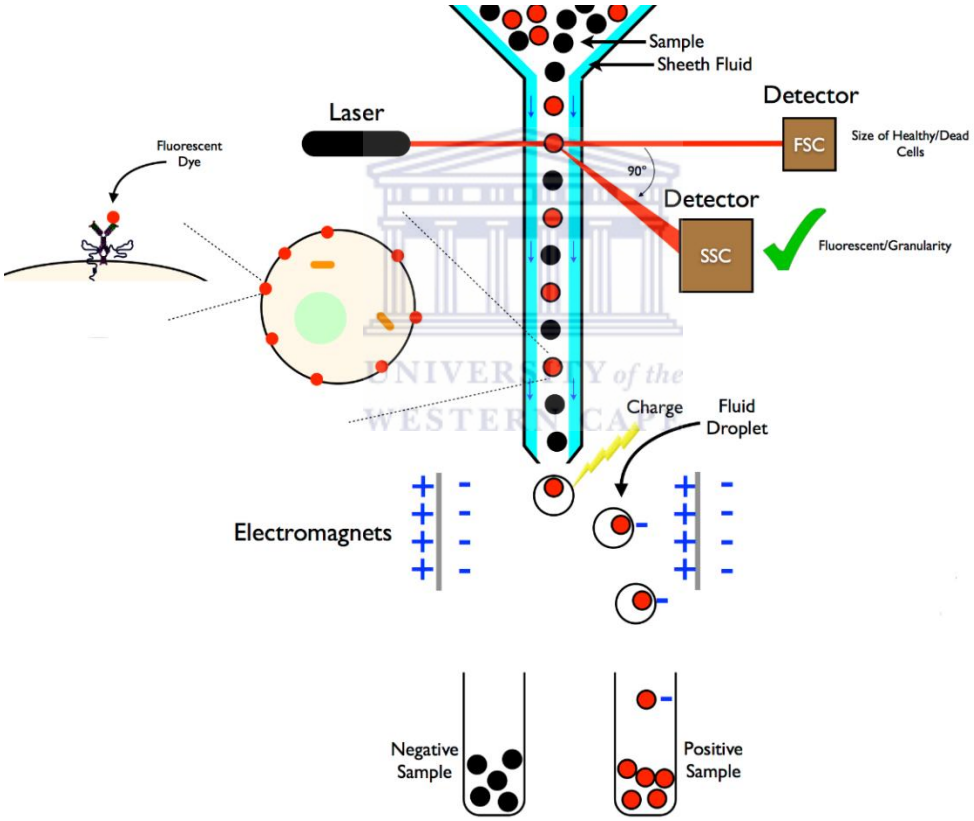

Figure 4.1: Graphical representation of the FACSorting process (Sabban, 2011).

Sorting of droplets is achieved by applying an electric charge to the stream which causes each newly formed droplet to carry a charge. The charged droplet containing cell(s) of interest are deflected into collection tubes using charged electrodes and are referred to as events. Other

droplets go to a waste stream. Parameters such as flow rate, number of events per second, break off point, and size and fluorescence intensity cut off can be adjusted where high throughput screening is required (Traganos, 1984; Herzenberg *et al*., 2002; Petersen & van den Engh, 2003). FACS data is analyzed using software and events are usually represented using dot plots or histograms (Figure 4.2).

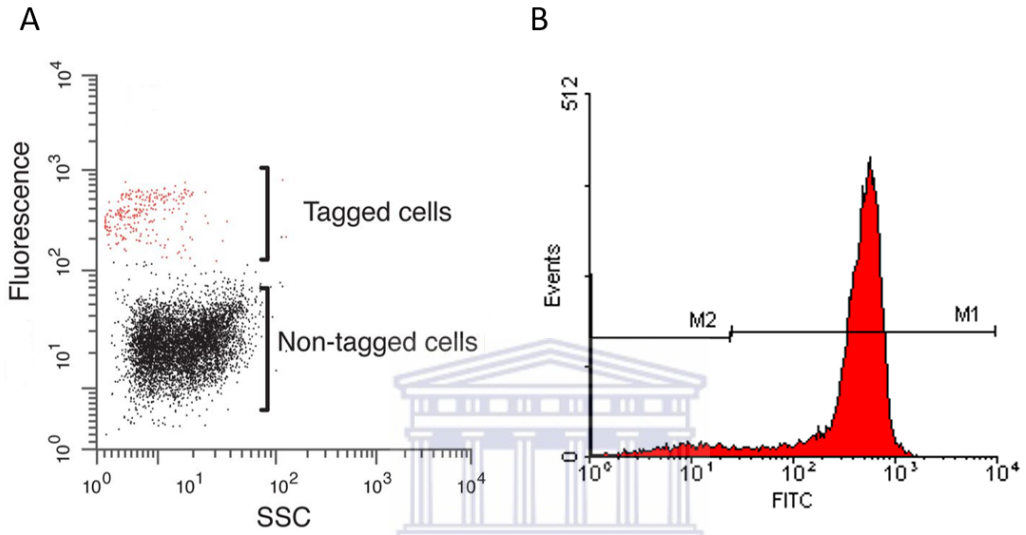

Figure 4.2*:* FACS data representation. (A) Dot plot of fluorescently tagged cells (Deng *et al*., 2012) (B) Histogram representing the fluorescence intensity of sorted events (Thomas, 2013).

Flow cytometry has been applied to viral studies. Examples include in the detection of viruses (Brussaard *et al*., 2000), development of procedures for counting viruses (Brussaard, 2004), detection of pre-inoculated bacterial hosts within mixed populations (Deng *et al*., 2012) and characterization of the infection dynamics of a novel model phage–host system (Allers *et al*., 2013). However, these studies have a number of limitations. First, they have been limited to known and well characterized viruses and bacterial hosts. Second, none of the procedures involved high throughput identification of bacteriophages and bacterial hosts.

4.1.2 Viral tagging

Viral tagging involves labelling phages with a fluorescent signal and challenging potential bacterial hosts with the tagged phages (Figure 4.3). The approach is high throughput, depending on no prior knowledge of the phage or host, and could potentially be used to isolate multiple phage-host pairs. Different approaches have been explored in generating phages with fluorescent tags. The protein coat of phages is generally permeable to many dyes and these are known to bind to nucleic acids packaged within the phage head. Some studies have used nucleic acid intercalating dyes such as YOYO-1 for detecting *E. coli* O157:H7 in broth (Goodridge *et al*., 1999a) and SYBR gold- labelled phage *P22* for detecting *Salmonella typhimurium* LT2 in a bacterial mix (Mosier-Boss *et al*., 2003). Aromatic diamidines like 4‘,6-diamidine-2-phenylindole dihydrochloride (DAPI) have been shown to bind non-intercalatively to DNA and can also be used (Zimmer & Wahnert, 1986).

Figure 4.3*:* An overview of viral tagging. Fluorescently labelled viruses are incubated with potential hosts. Fluorescent bacterial cells indicate phage attachment, which can be detected by microscopy and/or flow cytometry (Deng *et al*., 2012).

Another approach has been to genetically modify the phage of interest to express a fluorescent protein on its coat. The gene encoding the fluorescent protein is expressed in newly assembled phage particles after host infection. An example of such genes is luciferase which was used for detecting *Mycobacterium tuberculosis*, emitting light on phage infection (Jacobs *et al*., 1993).

Other examples include the green fluorescent protein (GFP) used in *E. coli* systems (Funatsu *et al*., 2002; Tanji *et al*., 2004) and coliphage T7 which was engineered to express biotin fused to its major capsid protein gp10A (Edgar *et al*., 2006).

4.1.3 Fluorescent *in situ* hybridization

Fluorescence *in situ* hybridization (FISH) is a widely used technique for gene detection and involves hybridizing a nucleic acid probe to the complementary target within a cell (Strandberg & Enfors, 1991; Moter & Göbel, 2000; Thurnheer *et al*., 2004; Machado *et al*., 2013). It is typically used in phylogenetic studies for identifying specific bacterial species in a mixed population. For phylogenetic analysis, probes are designed to the 16S rRNA gene, chemically synthesized, single stranded and about 15-50 base pairs in length (Dernburg *et al*., 1996; O'Keefe *et al*., 1996; Femino, 1998). The 16S rRNA gene is abundant in cells, providing enough targets for the probe to hybridize to. Also, the probes, being relatively short, transverse the cell membrane easily. This eliminates the need for additional permeabilization treatments, and unlike double stranded probes, steps to keep complementary probe strands apart, e.g. snap cooling, are not required. A routinely used probe is the bacterial domain- specific EUB338 (Amann *et al*., 1990) although new variants, EUB338-II and EUB338-III, have been designed to allow for more accurate detection (Daims *et al*., 1999).

While FISH is mostly used for hybridization to RNA due to the abundance of targets, the method can also be applied to DNA. A recent study detected the *amoA* gene, which encodes the alpha subunit of the enzyme ammonia monooxygenase and mediates ammonia oxidation in *Crenarchaeota*, in Namibian seawater samples using a technique termed geneFISH (Moraru *et al*., 2010). This technique involved designing double stranded DNA probes to the *amoA* gene and hybridizing them to the complementary regions on the *Archaea* via a series of treatments.

The technique was also applied to document the infection dynamics between phage PSA-HP1 and its *Pseudoalteromonas* host and referred to as phageFISH (Allers *et al*., 2013). However, the study relied on previous characterization of the phage-host system which required culturing of the bacteria. Also, sampling was carried out periodically, ranging from the time the cells were infected with phages to the exit of newly assembled viruses, in an attempt to track the assembly and release of new phage particles, hence there were multiple copies of the phage genome within the bacterial cells at the later stages of the experiment. This provided multiple targets for probe hybridization, a scenario less likely to exist in environmental samples. Also, while probes of lengths between 300-800 base pairs have been found to be ideal (Allers *et al*., 2013), it becomes necessary, among other optimizations, to find the best permeabilization treatment that allows probe entry into the cell without compromising its integrity.

Also, in terms of isolating novel phage-host pairs from environmental fractions, the relationship between the phage and host is completely unknown, hence FISH is limited to the detection of lysogenic phages within the genomes of bacterial hosts. Lysogens generally exist as single copies within bacterial cells (Powell *et al*., 1995; Svenningsen & Semsey, 2014), therefore, FISH requires careful optimization to detect such low copy number species. One such optimization has been to improve the signal generated by each probe molecule.

A number of methods to improve probe signal exist. For example, fluorescently labelled dNTPs such as CY3-dCTP have been incorporated in to the 3' end of probes via an enzymatic reaction (Pogliano *et al*., 2001). Another method has been to incorporate a hapten or antigen such as dioxygenin (DIG) into probe synthesis using PCR (Moraru *et al*., 2010) (Figure 4.4).

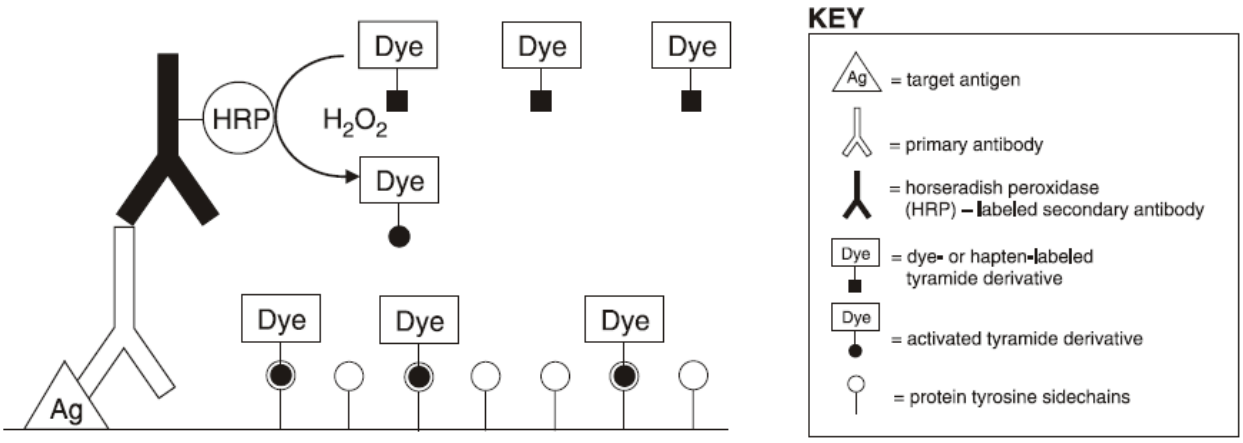

Figure 4.4: A graphical representation of the basic principle of tyramide deposition (www.mobitec.de/probes/docs/sections/0602.pdf) [Accessed: 07-06-2016].

The DIG molecule is incorporated every 4-5 bases, significantly amplifying the signal, and is detected by conjugating with antiDIG-horseradish peroxidase and followed by the deposition of fluorescently labelled tyramide in a series of chemical reactions, a process referred to as Catalysed reporter deposition (CARD) (Pernthaler *et al*., 2002; Amann & Fuchs, 2008; Clutter *et al*., 2010). Tyramide deposition is particularly useful in further amplifying the signal. In the presence of hydrogen peroxide, peroxidases cleave and activate the labelled tyramide causing a deposition of multiple copies around tyrosine side chains (de Haas *et al*., 1996; Macechko *et al*., 1997; Clutter *et al*., 2010). This helps to avoid diffusion related signal loss. However, the inactivation of endogenous peroxidases with cells is required to reduce background fluorescence. This study therefore aimed to develop a novel method for isolating phage host pairs from extreme environments using both viral tagging and FISH.

4.2 MATERIALS AND METHODS

4.2.1 Propagation of *Geobacillus thermoglucosidasius* 11955 and its phage, GVE3

The bacteria was routinely grown in TGP medium containing the following per litre: 17 g Tryptone, 3 g Soy peptone, 2.5 g K_2HPO_4, 5 g NaCl with the addition of filter sterilized 4 g Sodium pyruvate and 4 ml glycerol after autoclaving. The culture was grown at 60^{o}C with shaking at 250 rpm. Phage GVE3 was obtained from previous lab stocks. Viral lysates were prepared by infecting a 200 ml *G. thermoglucosidasius* culture at OD_{600} 0.4 with 100 µl (10^7 phages/ml) and incubating until lysis was observed. The culture was further clarified by adding 50 µl of chloroform and incubating for 15 minutes followed by centrifugation at 7000 rpm for 10 minutes. The supernatant was pushed through a 0.2 µm filter to exclude cellular debris. The filtrate was centrifuged at 19,000 rpm (Beckman J-26 XP, JA-20 rotor) for 6 hours. The supernatant was discarded and the translucent phage pellet resuspended in TE buffer.

The resuspended phage pellet was further purified by concentrating in a 3000 Da cut-off Amicon® Ultra-0.5 centrifugal filter (Merck Millipore) by centrifugation at 10,000 rpm for 10 minutes. The column was reversed and spun to collect phage particles which were re-suspended in TE buffer.

4.2.2 Preparation of environmental samples

The environmental phage and bacterial fractions were prepared as described in Section 2.2.1.

4.2.3 Biotin labelling of phage fraction

GVE3 phage (protein) concentration (mg/ml) was estimated using the Bradford method (Bradford, 1976). A 10 mM stock solution of Biotin (Life Technologies) was prepared by dissolving 2 mg of the reagent in 360 µl of ultrapure water. The amount of biotin required to

sufficiently label phage particles was estimated according to the following equations provided by the manufacturer:

i. ml protein X phage conc. (mg/ml) X $\frac{20\text{mmol}}{\text{M.W. (mg) of phage}}$ = mmol of biotin

ii. mmol of biotin X $\frac{1{,}000{,}000\ \mu l}{10\ \text{mmol}}$ = Volume of biotin reagent in µl

The calculated volume of biotin was added to 1ml of phage suspension and incubated at room temperature for 30 minutes or on ice for 2 hours.

4.2.4 SYBR gold labelling of phage fraction

DNase I (18 µl of 1 mg/ml) (Fermentas) and RNase A (8 µl of 12.5 mg/ml) (Fermentas) were added to 1.8 ml aliquot of GVE3 phage suspension and incubated at 37^{o}C for 30 minutes to digest free floating nucleic acids. Viral particles were stained by adding 0.2 µm- filtered SYBR gold (Life Technologies) to the suspension at 5X final concentration for concentrated viral stock (≥ E+09) or 1X for less concentrated viral stock (<E+09). This was carried out in lowered light. A negative control tube containing TE buffer without viruses was included as well to assess the efficiency of washing of excess SYBR Gold. The samples were mixed by gentle pipetting and the dye and viruses incubated in the dark at room temperature for 10 minutes followed by 80^{o}C for 10 minutes. The samples were allowed to cool for a further 10 minutes before use to avoid disruption of phage particles. To remove excess SYBR gold stain, Amicon® Ultra-0.5 centrifugal tubes (Merck Millipore) were used as follows: the walls of centrifugal tubes were initially blocked to prevent phage adhesion by adding 500 µl 0.02 µm filtered TE buffer, allowing to stand for 10 min, and spinning until almost dry (5,000 rpm, 10 minutes). The tubes were blocked by adding 500 µl of freshly prepared, 0.2 µm filtered 1%

BSA (10 mg/ml), allowed to stand one hour at room temperature and spun until almost dry (5,000 rpm, 15-30 minutes). This helped prevent adhesion of phage particles to the walls of the tubes device. The tubes were washed once with 500 µl of 0.2 µm filtered TE buffer and centrifuged at 5,000 rpm for 10 minutes.

Fluorescently labelled viruses and the negative control were concentrated in the centrifugal tubes at 3000 rpm for 15-30 minutes at 10^{o}C to a final volume to less than 50 µl. The reduced volume of stained virus or virus-blank was washed with 500 µl of 0.2 µm filtered TE buffer and centrifuged at 5000 rpm for 15-30 minutes. Washes were repeated a total of 6 times making sure to bring the volume to <50 µl each time. The washed viruses were transferred into fresh tubes using a pipette. To detach viral particles which may have remained "stuck" to the filter, 50 µl of 0.2 µm filtered TE buffer was added to each tube and sonicated for 3 min at 50W and 42 kHz. The suspension was pipetted up and down on the filter carefully so as not to puncture the membrane and transferred to the appropriate stained virus or blank tubes. Alternatively, labelled viruses were washed by dialyzing against PBS buffer. Washing efficiency was checked by adding non GVE3 infecting *E. coli* cells to a fraction of the phage suspension, incubating for 10 minutes and observing under a fluorescent microscope. The cells should remain unstained. Viral tagging was performed as soon as possible; else samples were stored at 4^{o}C.

4.2.5 Preparation of bacterial fraction for viral tagging

One ml of *G. thermoglucosidasius* 11955 cells at OD_{600} 0.4 was pelleted at 5000 rpm for 10 minutes, washed twice with, and re-suspended in, 500 µl of 1x PBS (1370 mM NaCl, 27 mM KCl, 80 mM Na_2HPO_4, and 20 mM KH_2PO_4). Biotin or similar molecules on the bacterial cell surface were blocked by adding 50 µl of stock streptavidin solution (1mg/ml) (Life Technologies) and incubating for 10 minutes. The bacterial cells were washed twice with 1x

PBS (5000 rpm, 5 minutes) to remove unreacted streptavidin and re-suspended in 1 ml of the buffer.

4.2.6 Phage- bacterial interaction

One µl of Streptavidin-FITC conjugate (Life Technologies) was added to 100 µl of biotin labelled phage in a separate 1.5 ml Eppendorf tube and incubated for 15mins. The biotin-streptavidin-FITC labelled phage was added to the 1ml streptavidin blocked cells and incubated for 20 minutes at room temperature to allow phages to bind to host cells. The whole mixture was centrifuged at 5000 rpm for 10 minutes to collect cells which should have fluorescently labelled phages attached to them. The cells were washed twice with 1X PBS to eliminate non-reacted streptavidin-FITC conjugate, re-suspended in 200 µl PBS and stored at 4^{o}C until viewed under the fluorescence microscope.

To stain the bacterial fractions, one millilitre of each bacterial sample (*G. thermoglucosidasius* 11955 culture at O.D.$_{600}$ 0.4, for control experiments, and the environmental bacterial fraction) was centrifuged at 5000 rpm for 10 minutes, washed twice and re-suspended in 200 µl of TE buffer. One microlitre of ethidium bromide (10 mg/ml) was added and the cell suspensions incubated for 5 minutes followed by three washes with TE buffer at 500 rpm for 5 minutes. Fifty microlitres of the SYBR gold labelled phage suspension was added to 200 µl of cells and incubated at room temperature for 20 minutes to allow for phage adsorption. The mixture was centrifuged at 5000 rpm for 10 minutes, washed twice, re-suspended in TE buffer and observed under a fluorescence microscope.

4.2.7 PCR amplification of FISH probes and target

Probes and target used in this study are summarized in Table 4.1 below. The reactions were performed in a thermal cycler Gene Amp PCR system. Each 50 µl reaction was carried out in 0.2 ml thin walled tubes containing 1X Phusion buffer, 200 µM each of dNTPs, 0.5 µM of each primer, 200 ng of template DNA and 0.02U/µl of Phusion DNA polymerase (Thermo Scientific).

Table 4.1: A summary of the probes and targets used in this study.

Probe	Length	Target	Primers	PCR amplification conditions
DS	300 bp	Contig 179 from the saline site metavirome (Section 2.3.3.2)	DS300F1: 5'ATCGGTACTCGCTACAGGTT-3' DS300R1: 5'-CCCAAACCCCCATCAGAAAA-3'	1 cycle of 98°C for 3 minutes; 34 cycles of 98°C for 10 seconds, 55°C for 30 seconds, and 72°C for 15 seconds; and 72°C for 10 minutes; and holding at 4°C.
MHB	300 bp	*Salinivibrio*--infecting phage SMHB1 (Chapter three)	MHB1PF: 5'-ACGCTACTCGAAACACCGGA-3' MHB1PR: 5'- GTTCTCGCGCATCACCTTTT-3'	1 cycle of 98°C for 3 minutes; 34 cycles of 98°C for 10 seconds, 60°C for 30 seconds, and 72°C for 15 seconds; and 72°C for 10 minutes; and holding at 4°C.
DS5	4.8kb	Contig 179 from the saline site metavirome (Section 2.3.3.2)	DS5For: 5'-ACACTCGGAAGCCATCATCTG-3' DS5Rev: 5'-CCCAGGTCGTAAGTGTGAGT-3'	1 cycle of 98°C for 3 minutes; 34 cycles of 98°C for 10 seconds, 55°C for 30 seconds, and 72°C for 2.5 minutes; and 72°C for 10 minutes; and holding at 4°C.

The PCR products were analyzed for the correct size by gel electrophoresis (Section 2.2.3.2) then purified using the NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel). The PCR products were sequenced at the Central Analytical Facility of Stellenbosch University, South Africa, to confirm that the correct fragment was amplified.

4.2.8 CY3-dCTP incorporation via PCR

CY3-dCTP was incorporated into the DS5 DNA fragment (Table 4.1) using DreamTaq DNA polymerase (Fermentas), Takara ExTaq DNA polymerase (Clontech) and Phusion DNA polymerase (Thermo Scientific). The reactions were carried out using primers DS5For (5'-ACACTCGGAAGCCATCATCTG-3') and DS5Rev (5'-CCCAGGTCGTAAGTGTGAGT-3') according to manufacturer's instructions. The ratio of dCTP to Cy3-dCTP in the dNTP mix was adjusted to 3:2. Controls included reactions with no CY3-dCTP. The thermal cycling conditions are summarized in Table 4.2 below:

Table 4.2: Thermo-cycling conditions for DNA polymerases used for CY3-dCTP incorporation

DNA Polymerase	Cycling Conditions
DreamTaq	1 cycle of 95°C for 3 minutes; 34 cycles of 95°C for 10 seconds, 62°C for 30 seconds, and 72°C for 30 seconds; and 72°C for 10 minutes; hold at 4°C.
Takara ExTaq	1 cycle of 95°C for 3 minutes; 34 cycles of 95°C for 10 seconds, 62°C for 30 seconds, and 72°C for 30 seconds; and 72°C for 10 minutes; hold at 4°C.
Phusion	1 cycle of 98°C for 3 minutes; 34 cycles of 98°C for 10 seconds, 62°C for 30 seconds, and 72°C for 15 seconds; and 72°C for 10 minutes; hold at 4°C.

CY3-dCTP incorporation using phi29 DNA polymerase was performed using Illustra GenomiPhi HY DNA Amplification Kit (GE Healthcare). Briefly, 2.5 µl (10 ng) of probe DNA was added to 22.5 µl of sample buffer, denatured at 95ºC for 3 minutes and snap cooled on ice to keep the probe single stranded. A master mix containing 20.5 µl of reaction buffer, 2 µl (200 pmol) of CY3-dCTP, and 2.5 µl of enzyme mix was prepared on ice. The master mix was transferred to the cooled sample on ice and the reaction was incubated at 30ºC for 4 hours followed by inactivation of the DNA polymerase at 65ºC for 10 minutes. The labelled probe was precipitated overnight at 4ºC by adding 2 volumes of absolute ethanol and 1/10th volume of 3M sodium acetate (pH 5.2), followed by pelleting at 13000 rpm for 10 minutes. The DNA pellet was washed once with 70% Ethanol and resuspended in 50 µl of TE buffer.

Equal amounts of the PCR reactions were loaded on a 1% agarose gel to comparatively check the efficiency of CY3-dCTP incorporation.

4.2.9 3’ end Cy3-dCTP incorporation

One microgram of the DS5 DNA fragment was digested using *Alu*I, *Hha*I, *Mse*I and *Tai*I to generate smaller fragments with 3’ overhangs (Section 3.2.5.1). Complete digestion was checked on a 2 % agarose gel (Section 2.2.3.2) before pooling and cleaning up using the NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel). The smaller fragments were labelled with CY3-dCTP at the 3’ end using Terminal Deoxynucleotidyl Transferase, Recombinant (rTdT) (Promega) according to a modified version of the manufacturer’s recommendation. Briefly, stock CY3-dCTP was diluted to 100 pmol using the recommended dilution buffer containing 50 mM potassium phosphate, 100 mM NaCl, 1 mM β-mercaptoethanol, 0.1% Tween 20 and 50% glycerol. Each 50 µl labelling reaction contained the following: 1X Terminal Transferase Buffer, 350 ng of target DNA, 600 pmol of CY3-

dCTP, 2 µl of rTdT and 17 µl of ultrapure water. The reaction mixture was incubated at 37ºC for 60 minutes followed by enzyme inactivation at 70ºC for 10 minutes. Labelled probe was purified using NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel) and eluted with nuclease free water. Probe concentration and Cy3-dCTP incorporation was measured using a NanoDrop ND-100 Spectrophotometer (NanoDrop Technologies).

4.2.10 Digoxigenin labelling

Digoxigenin (DIG) labelled DS and MHB probes (Table 4.1) were produced by incorporating DIG-dUTP into dsDNA *via* PCR (70 µM Dig-dUTP) using the PCR DIG Probe Synthesis Kit (Roche) according to the manufacturer's instructions. Probe specific primers and annealing temperatures were used. The PCR products were checked by electrophoresis on 2% agarose gels for correct size and DIG incorporation (Section 2.2.3.2). Gel bands were cut and purified using the NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel). Probe concentration was determined using a Qubit® 2.0 Fluorometer (Life Technologies) and stored at -20^{0}C.

4.2.11 Melting temperature analysis of DIG labelled probe in hybridization-like buffer

The melting temperature of the DS and MHB probes under hybridization conditions was determined with a melting curve analysis using a LightCycler® 480 System (Roche). The hybridization-like buffer (wash buffer I) contained 35% formamide, 5X saline sodium citrate (SSC), 20 mM EDTA, 10% dextran sulphate, and 0.1% SDS. One microlitre of SYBR gold (10, 000 X) and approximately 300 ng of probe DNA was added to 100 µl of the hybridization mixture and 25 µl of the resulting solution was aliquoted per well. The reaction included a mixture of unlabelled and DIG labelled probe (1:1). Control reactions included unlabelled probe only, DIG labelled probe only and a no DNA control. The thermal cycling protocol used to conduct the hybridization and melting analysis was as follows: denaturation at 80°C (up to

95°C) for 5 min, hybridization at 42°C for 25 min and melting from 50°C to 75°C, +4.4°C per second, minimum ramp rate.

Similar measurements were made for wash buffer II whose composition was 0.1x SSC, 0.1% SDS, 1 µl SYBR gold (10,000X) and dsDNA (~240 ng per 25 µl reaction). The thermal protocol used for T_m analysis in this wash buffer was: from 50°C to 75°C, +4.4°C per second, minimum ramp rate.

4.2.12 Preparation of chemically competent cells

A single colony from a fresh streak of *E. coli* BL21 was used to inoculate 5 ml of Luria Broth (per litre: 10g Tryptone, 5g Yeast extract, 10g NaCl) and incubated overnight at 37ºC with shaking at 150 rpm. One millilitre of the overnight culture was used to inoculate pre-warmed 50 ml Luria Broth and grown until $O.D._{600}$ 0.5. The cells were transferred to a 500 ml centrifuge tube, incubated on ice for 15 minutes and harvested by cold (4ºC) centrifugation at 4500 x g (Beckman J-26 XP, JA-10 rotor) for 5 minutes. The pellet was gently re-suspended in 10 ml of ice cold 0.1 M $MgCl_2$, incubated on ice for one hour and harvested by centrifugation as above. The pellet was again gently re-suspended in 1 ml of ice cold 0.1 M $CaCl_2$ containing 15 % glycerol. Aliquots of 50 µl were transferred into 1.5 ml microtubes (Eppendorf) and stored at -80ºC.

4.2.13 Ligation

The DS5 fragment to be used as FISH target was cloned into pJET 1.2/blunt cloning vector (Thermo Scientific) according to the manufacturer's instruction. Briefly, the ligation reaction contained 1X reaction buffer, 0.05 pmol pJET cloning vector, 1 µl T4 DNA ligase, 0.15 pmol probe DNA and the reaction volume was made up to 20 µl using nuclease- free water. The

ligation mixture was incubated in a thermal cycler Gene Amp PCR system at 22°C for 20 minutes and stored at -20°C until required.

4.2.14 Transformation

An aliquot of 5 µl of the ligation mixture was added to 50 µl of *E. coli* BL21 chemically competent cells. The mixture was incubated on ice for 30 minutes followed by heat shock at 37°C for 5 minutes. A 500 µl volume of Luria Broth was added to the cells and incubated at 37°C for 30 minutes. Fifty microliters of the transformation mix was aseptically spread on LB agar containing Ampicillin (200 µg/ml) and incubated for 16 hours at 37°C. Colony PCR (Section 2.2.3.1) and restriction digest (Section 3.2.5.1) was used to select for positive clones. A clone containing the *csaB* gene from phage-resistant *G. thermoglucosidasius* in the pJET vector, provided in the lab (van Zyl *et al*., 2016), was used as a negative control for FISH experiments. Glycerol stocks of clones were stored at -80°C.

4.2.15 Optimization of FISH experiment

The FISH experiment was optimized in an *E. coli* BL21-pJET system before testing on a mix of bacteria isolated from a salt playa in Namibia (Chapter two). FISH in solution was carried out as described below (Pernthaler *et al*., 2002; Moraru *et al*., 2010; Manti *et al*., 2011; Allers *et al*., 2013). A summary of the protocol is provided in Figure 4.4 below:

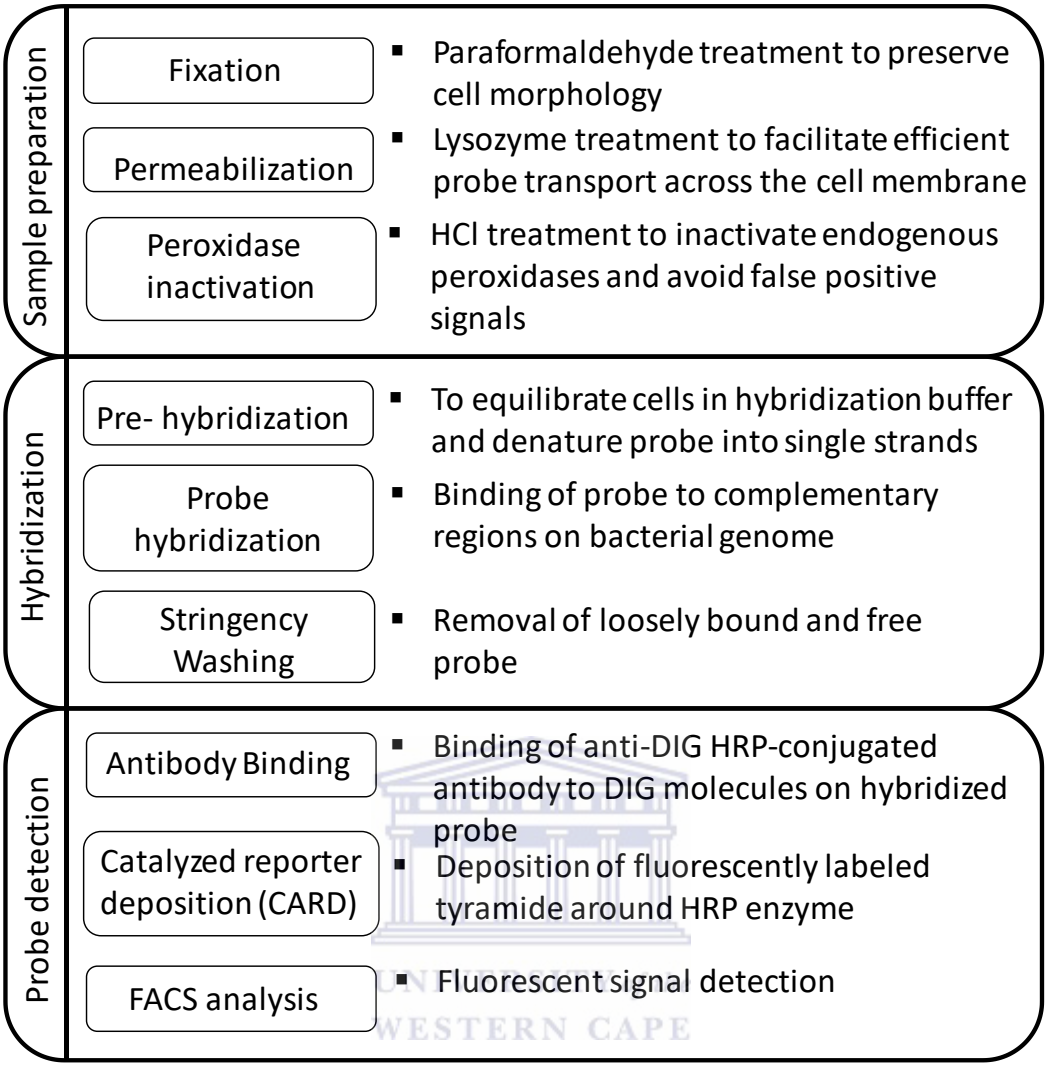

Figure 4.5: A summary of the FISH protocol used in this study.

4.2.15.1 Sample preparation

For control experiments, DS/pJET/BL21 clones containing the FISH target and negative control, *csaB*/pJET/BL21, were inoculated into 10 ml of Luria broth supplemented with Ampicillin (200 µg/ml), grown at 37^{o}C and harvested at O.D.$_{600}$ 0.6. Cultures were aliquoted such that each tube contained approximately 4.8 X 10^{8} cells and immediately fixed in paraformaldehyde.

A mock bacterial mix was also prepared using 8 isolates from the Swakop saline site described in Chapter two. Each isolate was grown to O.D.$_{600}$ 0.6 and the ratio of OD to viable cells

determined using the standard plate count method. An equal number of cells from each isolate was combined to make the mock environmental mix. Depending on the experimental set up, target cells (clones of *E. coli* or isolate SS3) were added to the mix in ratios 1:5 and 1:100. The target containing mixes were centrifuged at 5000 rpm for 5 minutes, washed twice with 1x PBS (5000 rpm, 5 minutes) and immediately fixed in paraformaldehyde.

4.2.15.2 Paraformaldehyde fixation

Two grams of paraformaldehyde (PFA) powder was dissolved in 50 ml of phosphate buffered saline (PBS; 130 mM NaCl, 10 mM Na_2HPO_4/NaH_2PO_4, pH 7.4) and heated at 60^oC for approximately 30 minutes until the suspension was clear. A few drops of 1M NaOH were added to aid complete dissolution and the pH was adjusted to 7. The solution was filtered through a 0.2 µm single unit filter (GVS Filter Technology) and placed on ice. The cells were washed twice with ice-cold PBS and re-suspended in 750 µl of the same buffer. The cell suspension was mixed with 250 µl PFA fixative and incubated at 4^oC overnight. After incubation, the cells were washed twice with ice cold PBS to remove as much fixative as possible and re-suspended in PBS and ice cold ethanol (1:1) then stored at -20^oC until required.

4.2.15.3 Lysozyme treatment

The fixed cells were permeabilized by re-suspending in a solution containing 0.5 mg ml^{-1} lysozyme, 1x PBS pH 7.4 (1370 mM NaCl, 27 mM KCl, 80 mM Na_2HPO_4, and 20 mM KH_2PO_4), 0.1 M Tris-HCl pH 8.0 and 0.05 M EDTA pH 8.0, for one hour, on ice. The cells were pelleted at 5000 rpm for 5 minutes, followed by re-suspension in 1x PBS pH 7.4 for 5 minutes. The cells were further washed by centrifuging (5000 rpm for 5 minutes) and

resuspending in distilled water for 1 minute. The cells were pelleted again (5000 rpm for 5 minutes) followed by the inactivation of endogenous peroxidases.

4.2.15.4 Inactivation of endogenous peroxidases

A number of methods were tested to select for one that sufficiently inactivated endogenous peroxidases while preserving cell morphology. These included (a) re-suspending cells in 3% hydrogen peroxide dissolved in either methanol or PBS (Streefkerk, 1972; Fink *et al*., 1979), (b) 0.01 M HCl and (c) 0.2 M HCl. After each treatment, the cells were incubated for 10 minutes, followed by washing with 1x PBS for 5 min, water for 1 min and 96% ethanol for 1 min, and by 5 minutes air-drying. The efficiency of each method was determined by adding Alexa $_{488}$-labelled tyramide to each sample, washing and observing under a fluorescence microscope.

4.2.15.5 Hybridization

Samples were pre-hybridized at the optimum temperature determined using the melting curve analysis (Section 4.2.11). Cells were re-suspended in 200 µl of hybridization buffer without probe for 3 hours. The hybridization buffer was either sodium chloride or saline- sodium citrate (SSC) based. Components of the hybridization buffers are summarized below:

Table 4.3: Sodium chloride based hybridization buffer (Fuchs *et al*., 2010).

Stock Reagent	Volume	Final Concentration in Hybridization Buffer
5 M NaCl	360 µl	900 mM
1 M Tris –HCl	40 µl	20 mM
Formamide	% depending on probe	
Distilled Water	Add up to 2 ml	
10% SDS	2 µl	0.01%

Table 4.4: Saline-sodium citrate based hybridization buffer (Moraru *et al*., 2010).

Component	Volume/Amount	Final Concentration
100% Formamide (FA)	14 ml	35%
20x SSC	10 ml	5X
Dextran sulfate (powder)	4 g	10%
20% SDS	200 µl	0.1%
0.5 M EDTA	1.6 ml	20 mM
miliQ Water	4.4 ml	
Sheared salmon sperm DNA (10 mg/ml)	1 ml	0.25 mg/ml
Yeast RNA (10 mg/ml)	1 ml	0.25 mg/ml
10% Blocking reagent for nucleic acids	4 ml	1%
Final volume	40 ml	

Probes were added at a final concentration of 300 pg/µl. Unlabelled probe was added to samples as negative control to check the efficiency of antibody and labelled tyramide wash steps. Other controls were determined based on the experiment. Each sample was incubated at 85^{o}C for one hour to sufficiently denature the probe and the bacterial genome. Standard FISH experiments were conducted using the recommended NaCl based hybridization buffer with incubation at 46°C for 18-22 hours (Moter & Göbel, 2000; Pernthaler *et al*., 2002; Fuchs *et al*., 2010). For hybridizations using SSC based buffer, samples were incubated at the pre-determined optimum hybridization for 18-22 hours (Allers *et al*., 2013).

4.2.15.6 Stringency washes

Stringency washes were carried out to remove unbound or loosely bound probe (Table 4.5 and Table 4.6).

For NaCl based hybridization, 1 ml of pre-warmed (48^{o}C) wash buffer was added to each sample prior to centrifugation at 5000 rpm for 5 minutes. Another 1 ml of wash buffer was added to each microtube and incubated at 48^{o}C for 10 minutes. The cells were centrifuged at 5000 rpm for 5 minutes, washed twice with 1X PBS, re-suspended in 50 µl of 1X PBS and stored at -20^{o}C until required. Components of the wash buffer are summarized below:

Table 4.5: Sodium chloride based standard wash buffer (Pernthaler *et al*., 2002).

Stock Reagent	Volume	Final Concentration in Hybridization Buffer
5 M NaCl	Concentration dependent on the amount of formamide used	See **Table 4.6**
1 M Tris –HCl	1 ml	20 mM
0.5 M EDTA (only if ≥ 20% formamide is used for hybridization	500 µl	5 mM
Distilled Water	Add up to 50 ml	
10% SDS	50 µl	0.01%

Table 4.6: Sodium chloride concentration in the washing buffer for washing at 48°C after hybridizing at 46ºC (Pernthaler *et al*., 2002).

Formamide in Hybridization Buffer (%)	[NaCl] in M Final Concentration	Microlitres of 5 M NaCl required in 50 ml
0	0.900	9000
5	0.636	6300
10	0.450	4500
15	0.3180	3180
20	0.225	2150
25	0.159	1490
30	0.112	1020
35	0.080	700
40	0.056	460
45	0.040	300
50	0.028	180
55	0.020	100
60	0.014	40
65	-	-
70	-	-

For SSC based hybridizations, samples were washed three times with wash buffer I (WBI; 2x SSC, 0.1% SDS) and incubated at the pre-determined optimum wash temperature for 30 minutes. This was followed by washing three times with wash buffer II (WBII; 0.1X SSC, 0.1% SDS) and incubation for 1.5 hours at the wash temperature in a slow shaking water bath. Samples were spun each time in a centrifuge kept at 40°C to avoid non-specific hybridization associated with lower temperatures. Cells were finally washed once with 1x PBS.

4.2.15.7 Antibody binding

Antibody binding took place in a solution containing 1x PBS, 1% blocking agent (Roche) and 0.3 U ml^{-1} (500x dilution of the 150 U/ml stock) anti-DIG HRP-conjugated antibody (Fab fragments; Roche) for 1.5 hours. Samples were washed three times in a solution of 1x PBS and 1% blocking agent for 10 minutes each. All steps were carried out with slow shaking.

4.2.15.8 Catalyzed reporter deposition (CARD)

Samples were re-suspended in 100 µl of amplification buffer (Table 4.7), 10 µl of 100X H_2O_2 (100X H_2O_2: 1 µl of 30% H_2O_2 + 200 µl of 1X PBS) and 2 µl of $Alexa_{488}$-labelled tyramide (Life Technologies) and incubated at 37°C for 10 minutes.

Table 4.7: Components of the CARD amplification buffer (Moraru *et al*., 2010).

Component	Volume	Final Concentration
10x PBS pH 7.4	4 ml	1x
5 M NaCl	16 ml	2 M
10% Blocking Reagent for nucleic acid hybridizations (BR)	400 µl	0.1%
Dextran Sulfate	8 g	20%
milliQWater	15.6 ml	

Excess labelled tyramide was removed from cells by washing three times with 1X PBS at 46^{o}C with slow shaking. Samples were re-suspended in 1X PBS and stored at 4^{o}C until visualized or FACS sorted.

4.2.16 Fluorescence activated cell sorting (FACS)

For viral tagging, samples were analyzed using a Becton Dickinson FACSAria™ flow cytometer at the Flow Cytometry Unit, University of the Western Cape, South Africa and the FACS data were interpreted using the BD FACSDiva version 6.0 software. The analysis of FISH experiments was conducted using a Becton Dickinson Accuri™ C6 flow cytometer at the Biotechnology department, University of the Western Cape, South Africa, and the data analyzed using FlowJo software version 10. The flow cytometers were equipped with violet (405nm) and blue (488nm) lasers while the FACSAria™ had an extra red (633nm) laser.

Dot plots of various fluorescent channels were used to enumerate populations and sub populations. SYBR Gold, FITC and $Alexa_{488}$ dyes were analyzed in the FITC/ FL1 channel. Ethidium bromide was analyzed in the PerCP- Cy5-5 channel. An initial sort was always conducted to adjust the position of each quadrant and gate when necessary.

4.3 RESULTS AND DISCUSSION

4.3.1 VIRAL TAGGING

4.3.1.1 Biotin labelling

Biotinylation is widely used for detection and purification of proteins or other macromolecules. The process involves attaching biotin to specific functional groups or residues such as carbohydrates, primary amines, carboxyls and sulfhydryls on the surface of the molecule of interest (Inouye & Nakamura, 2003; Edgar *et al*., 2006). The attached biotin molecule is detected by conjugating with fluorescently labelled avidin or streptavidin (Gitlin *et al*., 1987). This principle was employed for phage labelling based on the assumption that the protein coat of phages would have at least one functional group or residue (a free amine group – any lysine residue or the protein N-terminal) displayed on its surface to which biotin can be covalently coupled. To assess this, *G. thermoglucosidasius* 11955 and its phage, GVE3 (van Zyl *et al*., 2015), were used to optimize the labelling protocol before testing on the environmental fractions. The *Geobacillus thermoglucosidasius* 11955/ phage GVE3 system was used to optimize the labelling protocol before testing on the environmental fractions as it has been extensively studied in our laboratory (Van Zyl *et al*., 2014; van Zyl *et al*., 2015; van Zyl *et al*., 2016). However, consistent biotin labelling of the control phage fraction could not be achieved. Microscopic observation showed that a few phages were biotin tagged which was low compared to the number of viruses originally present. Doubling the amount of biotin used did not improve labelling. At one instance where a significant proportion of the phages were successfully labelled (microscopic observation), upon addition to the bacterial fraction, the majority of the cells had no fluorescent tags (Figure 4.6). It is possible that the reaction was suboptimal at the biotin binding step or the streptavidin-FITC conjugation step. DNA fragments can be checked for biotin incorporation via gel electrophoresis because biotin

reduces the electrophoretic mobility (Holmstrøm & Rasmussen, 1990), but this is impractical for intact phages that are still required to attach to bacterial cells.

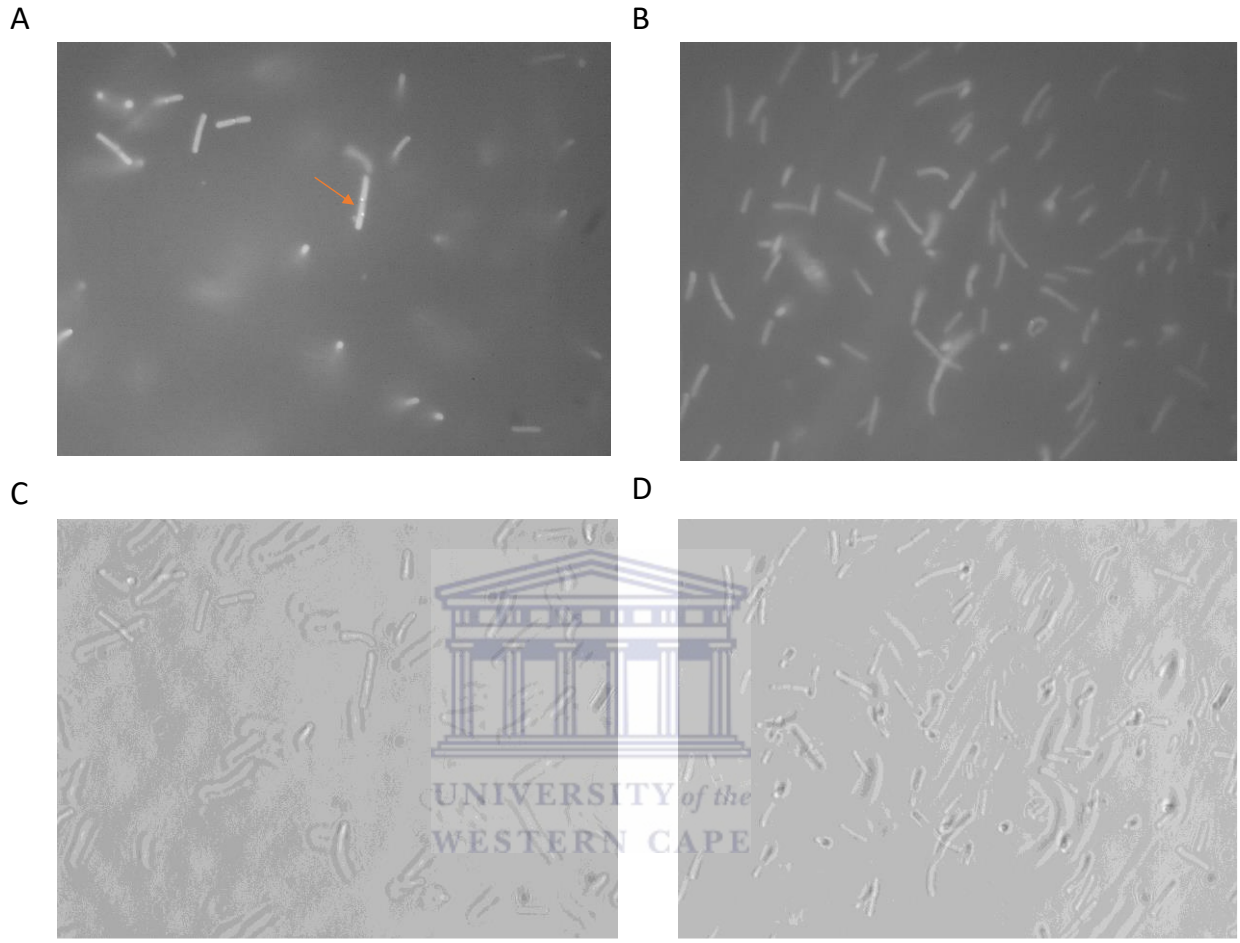

Figure 4.6: Biotin labelling of GVE3 phage particles A & C. *G. thermoglucosidasius* cells after the addition of biotin-streptavidin-FITC labelled GVE3 phages, under fluorescent (A) and bright field (C) views. The red arrow indicates probable attachment of labelled phage to cells. B & D. Negative control, with phages without biotin but streptavidin-FITC treated, under fluorescent (B) and bright field (D).

Also, biotinylation requires careful optimization such that the target protein has biotin attached at a minimal number of sites, ideally one, to avoid drastic changes to the surface properties of the protein (Koide *et al*., 2009). This becomes especially important in phage labelling where structural changes might negatively impact the ability of the labelled phage to attach to bacterial receptors. The inability of most of the GVE3 phages observed (Figure 4.6A) to bind

to the *Geobacillus* host may suggest some modification of the attachment site. Some studies avoided this problem by engineering the phage to display biotinylated peptides on the protein coat (Edgar *et al*., 2006; Chung *et al*., 2014). However, given that the aim of this experiment was to ultimately label environmental phage fractions that have not been previously studied, optimizing biotinylation for such a mixed population would be difficult, hence an alternative approach, involving nucleic acid labelling using SYBR gold, was explored.

4.3.1.2 SYBR gold labelling

G. *thermoglucosidasius* and phage GVE3 were again used as controls. Phage GVE3 was labelled with SYBR gold and the absence of excess dye verified by mixing with a non-susceptible host, *E. coli* cells and observing under a fluorescence microscope. The *E. coli* cells did not appear to fluoresce. Satisfied that there was no excess dye, tagging experiments with *G. thermoglucosidasius* cells were conducted and showed both free and bound phage particles (Figure 4.7). This result proved that tagging on the basis of the specificity of phage-host interaction was possible, at least for this phage-host system.

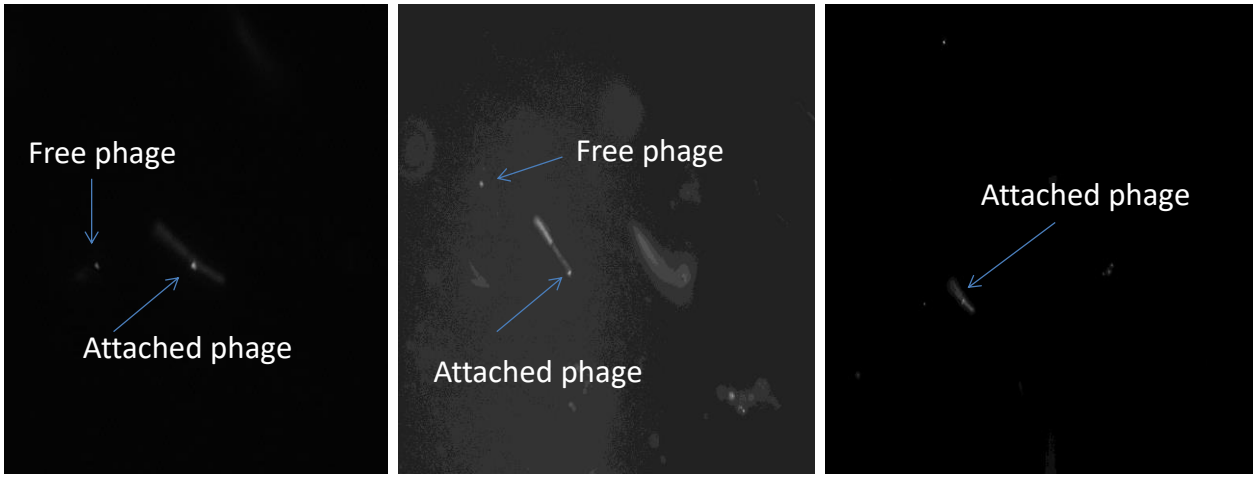

Figure 4.7: Micrographs showing SYBR gold labelled GVE3 phage particles attached to *G. thermoglucosidasius* cells. Each micrograph represents a different field of view.

Given that the aim of the study was to isolate novel phage-host pairs from environmental samples using flow cytometry and cell sorting, the control bacterial cells and phage particles were used to establish the initial parameters for where each population would be represented on a dot plot (Figure 4.8). A clear increase in fluorescence can be seen between ethidium bromide labelled (Figure 4.8A) and unlabelled (Figure 4.8B) *G. thermoglucosidasius* cells. Similarly, SYBR gold labelled GVE3 viruses (Figure 4.8C) were more fluorescent compared to the unlabelled fraction (Figure 4.8D). The few fluorescent events observed in Figure 4.8D were more likely to be background than inherent phage fluorescence.

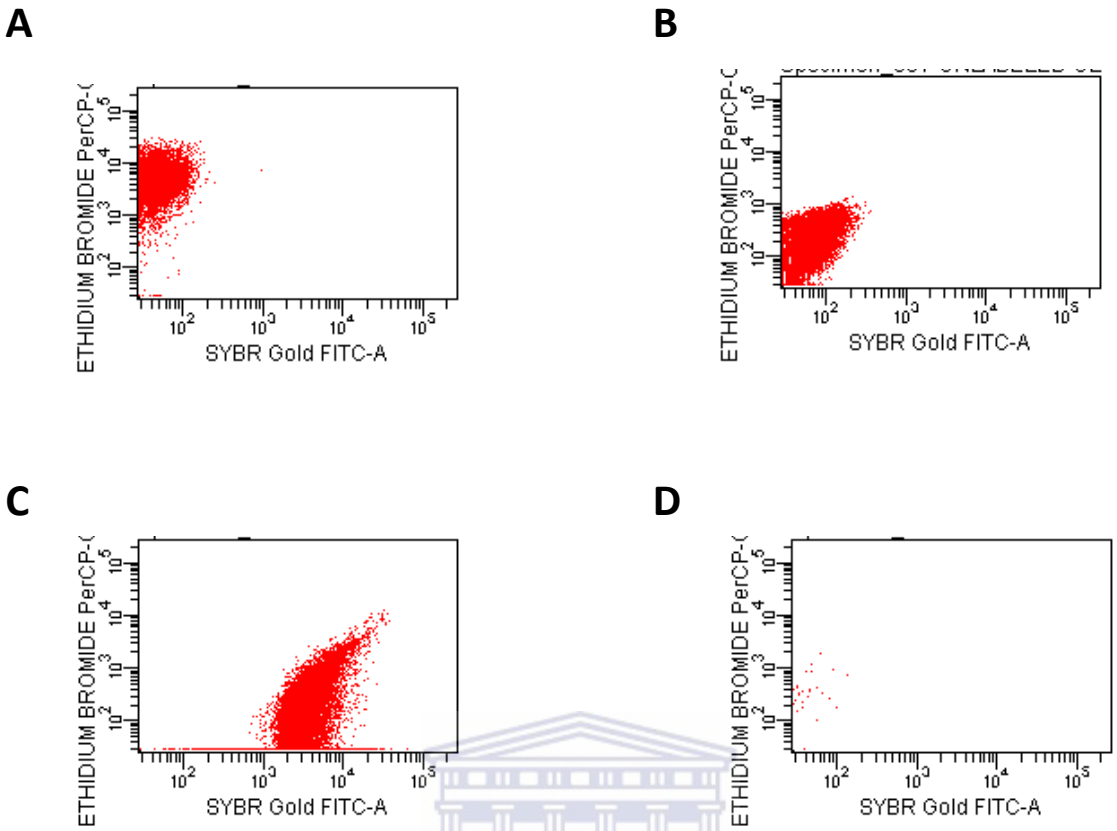

Figure 4.8: FACS dot plots showing the fluorescence intensity of labelled control events. Each plot represents one million events. A. *G. thermoglucosidasius* ethidium bromide labelled bacterial cells B. *G. thermoglucosidasius* unlabelled bacterial cells; C. SYBR gold labelled GVE3 phage suspension. D. Unlabelled GVE3 phage suspension.

Based on the preliminary results shown in Figure 4.8, gates and quadrants were introduced. EtBr-labelled *G. thermoglucosidasius* cells were represented as events in quadrant Q1 with increasing ethidium bromide fluorescence intensity along the y-axis. Events in quadrant Q4 represented SYBR gold labelled GVE3 phage particles. Events in Q3 represented bacterial cells or phage particles with low or no fluorescence. Gate P5 was designed to capture events that gave dual florescent signals which would be an ethidium bromide signal from the bacteria cell and a SYBR gold signal from the attached phage. This would represent a phage-host pair. The gate was adjusted to lie where no events were recorded when bacteria or phages were sorted individually.

The majority of the bacterial cells were observed to have an ethidium bromide fluorescence intensity above 10^3 (Figure 4.9A). The lack of complete separation of the population from the background could be attributed to some cells being in various stages of growth, hence, smaller cell volume which limited dye uptake.

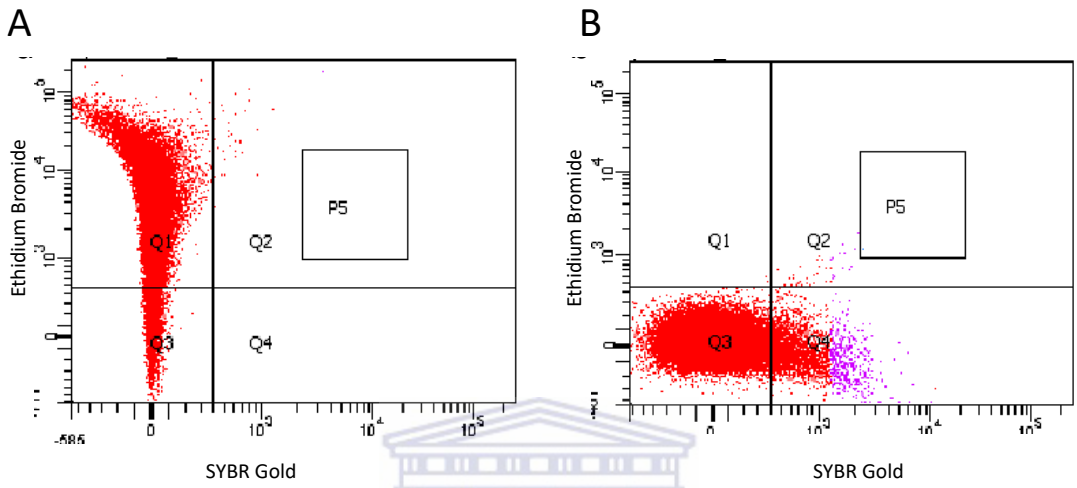

Figure 4.9: FACS dot plots showing the fluorescence intensity of labelled control events A: *G. thermoglucosidasius* ethidium bromide labelled bacterial cells; B: SYBR gold- labelled GVE3 phage suspension. Q1 represents bacteria cells with high fluorescence intensity. Q4 represents phages with high fluorescence intensity. Q3 contain events bacterial cells or phage particles with low or no fluorescence. Q2 contains gate P5 which should capture phage-host pairs.

Given that the control phage suspension was expected to be homogenous, it was expected that the fluorescent signal observed would be distinct and of high intensity, and distinctly separated from the unlabelled phage or debris. However, the broad range of signal observed for the phage particles (Figure 4.9B) suggested the presence of some excess dye which was not readily identified using the less sensitive fluorescence microscopy observation. The cell sorting capability of the flow cytometer was therefore explored as a way to selectively sort the labelled phage population from the background and excess dye, in essence using sorting as a wash step. FACSorting involves passing each event through a steady stream of PBS buffer just before interrogation. The samples were passed through the machine and collected in a new tube. This

enabled each cell and phage to be rinsed individually in a larger volume of PBS, thus removing excess dye. Given that the wash steps initially employed involved centrifugation, which may cause some phages to rupture, washing with FACSorting should eliminate this problem.

Another measure incorporated into the experimental design was to include additional gates to specifically select for a defined range of fluorescence. This would represent populations of similar sizes and fluorescence intensity. As shown in Figure 4.10A below, labelled bacterial cells were selected for using gate P2. The cells were re-sorted to confirm that the right population was selected for (Figure 4.10B).

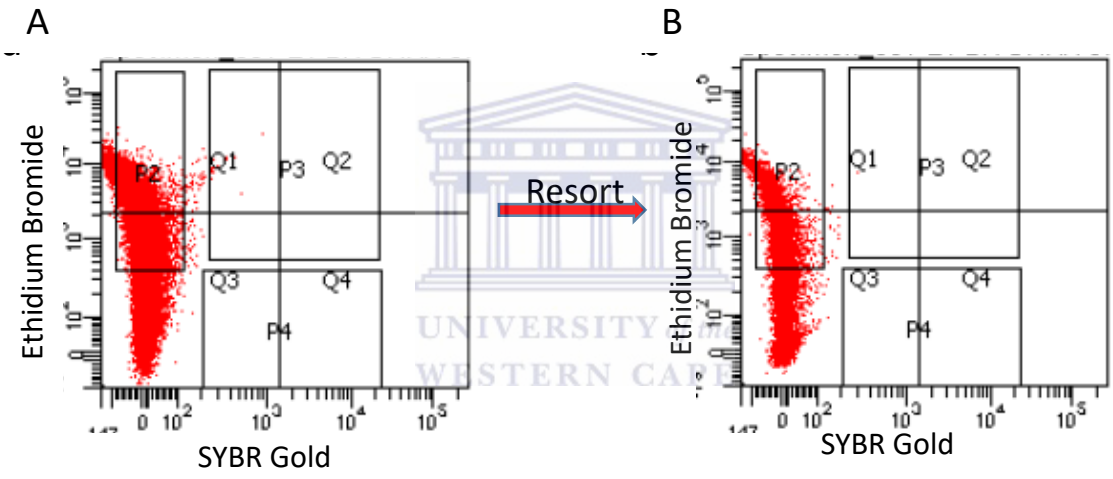

Figure 4.10: Dot plot of ethidium bromide-labelled *G. thermoglucosidasius* bacterial cells A: Initial sort B: Re-sort of the events collected in A.

However, the re-sort was not as accurate as expected as several events were recorded outside the defined gates. A similar result was observed for the GVE3 phage population (Figure 4.11).

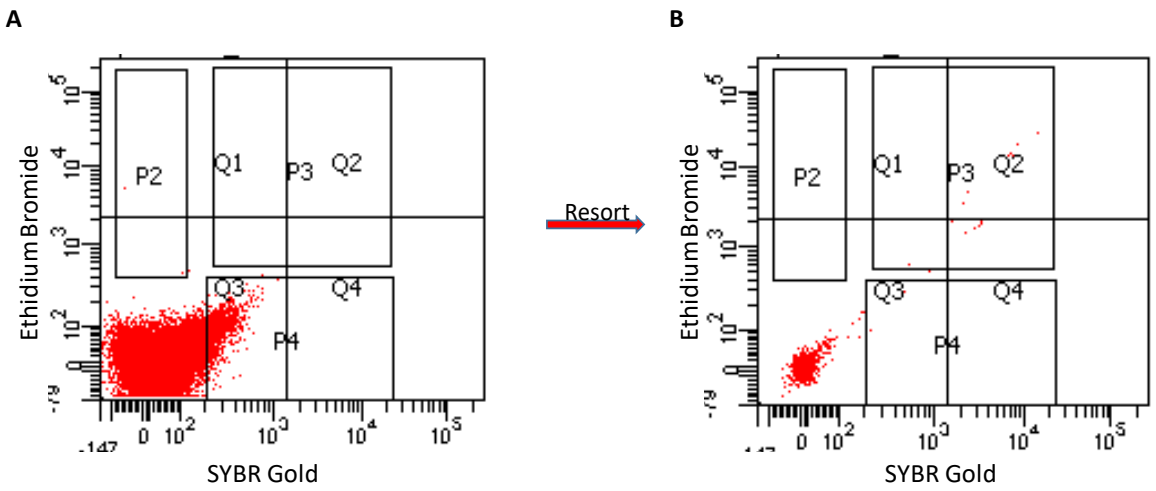

Figure 4.11: Dot plot of SYBR gold-labelled GVE3 phages A: Initial sort; B: Re-sort of the events collected in A.

It has been suggested that SYBR gold increases the pressure within the phage head which can trigger the premature mechanical ejection of the tightly packaged nucleic material in some phages (Ohno *et al.*, 2012). This may have been a reason for the lack of a distinct phage population (Figure 4.11A). A significant reduction in the number of phages was also observed while resorting (Figure 4.11B). This suggested that phages were lost probably due to the increased pressure mentioned above.

Another issue was the appearance of events in gate P3, where based on the fluorescence observed in Figure 4.10A, Figure 4.10B and Figure 4.11A, only dual labelled events should appear (Figure 4.11B). This raised a major question because the events sorted were SYBR gold labelled phages and suggested that ethidium bromide may also be carried over from previous experiments. The efficiency of the backwashing process was therefore checked using *E. coli* cells.

Sorting of *E. coli* cells after the recommended wash cycle (Figure 4.12A) and after a backwash (Figure 4.12B) showed similar profiles. Ethidium bromide labelled cells were also sorted after a backwash and aligned along the expected fluorescence axis (Figure 4.12C). However, a run of PBS buffer following the sorting of labelled cells and a backwash showed a significant number of unexpected events (Figure 4.12D). These events were not recorded in earlier experiments though the same PBS stream was applied. These events also coincided with the area of the dot plot where potential phage-host pairs should theoretically be found, thus increasing the chances of false positives. Changing the buffer and filtering through a 0.2 µm filter prior to sorting did not resolve the issue. This suggested a fundamental problem with the machine set up, which could not easily be resolved as the use of the FACSorter was provided as an external service.

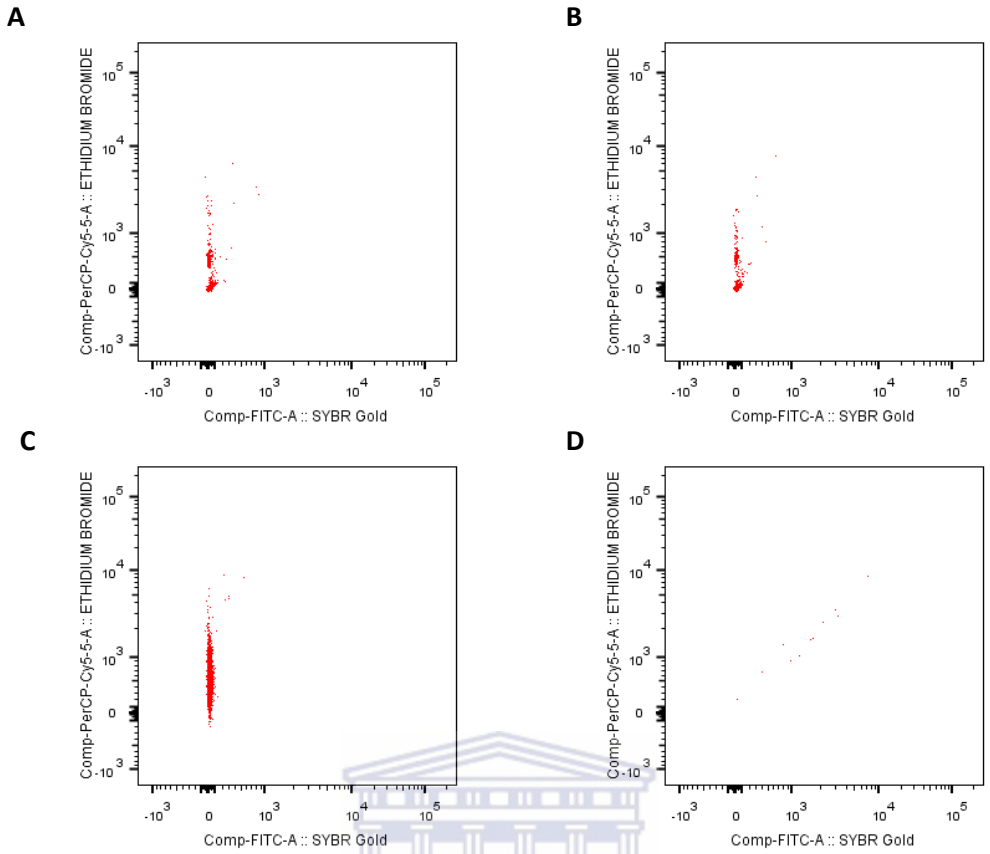

Figure 4.12: Dot plots of *E. coli* cells used to validate the backwashing process A. Plain cells pre-backwash B. Plain cells, post-backwash C. Ethidium bromide labelled cells, post backwash D. PBS buffer only, post-backwash.

A combination of the difficulty in removing excess dye, loss of phage particles and the appearance of questionable events in designated gates meant the attempt to isolate phage-host pairs *via* viral tagging was abandoned in favour of fluorescence *in situ* hybridization.

4.3.2 FLUORESCENCE *IN SITU* HYBRIDIZATION

4.3.2.1 Probe and target amplification from metaviromic DNA

Potential FISH targets were identified from phage contigs from the analysis of sequence data from the Swakop saline metavirome (discussed in Chapter two). One of the assembled contigs, Contig 179, was particularly interesting. The contig has a GC content of 38.5 % and 25 open reading frames. BLASTn analysis of the full length of Contig 179 showed 75% similarity to the cyanobacterium *Dactylococcopsis salina* PCC 8305 at 3% query cover and E-value of 7 E-79. However, BLASTx analysis of the ORFs (Table 4.8) show that the contig is most likely a prophage within the genome of the bacterium. Contig 179 was therefore selected for FISH experiments.

Table 4.8: A summary of the BLASTx analysis of predicted ORFs of contig 179. The + or – sign before the ORF number is indicative of the strand on which the gene is found.

ORF #	ORF Position	Closest Phage BLAST Hit	Identities	Positives	E-value	Accession Number
1/-	144-2516	DNA primase (Mediterranean phage uvMED)	97/373(26%)	182/373(77%)	1e-24	WP019504711.1
2/-	2823-3152	Hypothetical protein			0.0	
3/+	4322-4717	CI repressor protein (*Pseudomonas* phage PS-1)	17/55(31%)	32/55(58%)	0.019	YP009222798.1
4/-	4765-5172	Hypothetical protein			0.0	
5/-	5435-5767	Putative head morphogenesis domain protein (*Vibrio* phage SSP002)	13/41 (32%)	25/60(60%)	2.5	AFE86358.1
6/+	5985-6530	Hypothetical protein (Mediterranean phage uvMED)	12/27 (44%)	16/27 (59%)	8.9	BAQ85063.1
7/+	6677-6979	Hypothetical protein			0.0	
8/-	7096-8193	Integrase (Mediterranean phage uvMED)	87/327 (27%)	139/327 (42%)	5e-22	BAR30339.1
9/-	8536-9942	Hypothetical protein			0.0	
10/-	9942-10961	Hypothetical protein			0.0	

11/-	10952-12654	Putative DNA polymerase (*Klebsiella* phage JD001)	13/36 (36%)	19/36 (52%)	9.7	YP007392876.1
12/-	12563-13045	Hypothetical protein (Mediterranean phage uvMED	14/30 (47%)	18/30 (60%)	7.1	BAQ89139.1
13/-	12563-13045	Hypothetical protein			0.0	
14/-	12984-13646	Hypothetical protein			0.0	
15/-	13643-14449	Hypothetical protein			0.0	
16/-	14454-15185	Hypothetical protein			0.0	
17/-	15176-15526	Hypothetical protein			0.0	
18/+	16200-16967	YapH potein (Mediterranean phage uvMED)	11/19 (58%)	12/19 (63%)	2.3	BAQ85281.1
19/+	17012-17770	Hypothetical protein			0.0	

20/+	17770-18147	Hypothetical protein (*Aeromonas* phage CC2)	11/17 (65%)	13/17 (76%)	6.3	YP007010037.1
21/+	18128-18505	Hypothetical protein			0.0	
22/+	18465-19091	Hypothetical protein			0.0	
23/+	19093-19506	Hypothetical protein			0.0	
24/+	19506-19868	Tryptophan halogenase (*Synechococcus* phage S-ShM2)	11/34 (32%)	21/34 (17%)	0.001	YP004322694.1
25/+	19858-20250	Hypothetical protein			0.0	

The presence of genes such as the CI repressor protein (ORF 3) which helps to maintain latency (Aggarwal *et al*., 1988; Czyz *et al*., 2001) and the integrase gene (ORF 8) which facilitates lysogeny (Shotland *et al*., 1997; Parua *et al*., 2010) suggested that it might be a lysogenic phage. It was therefore selected for control FISH experiments.

A 4.8kbp (designated DS5) fragment was amplified and cloned for use as target in control experiments (Figure 4.13). The use of the 4.8kbp fragment to generate probes of random lengths was explored. Specific primers were also designed to amplify a 300 bp fragment of the 4.8 kb used as probe. The probe was designated DS.

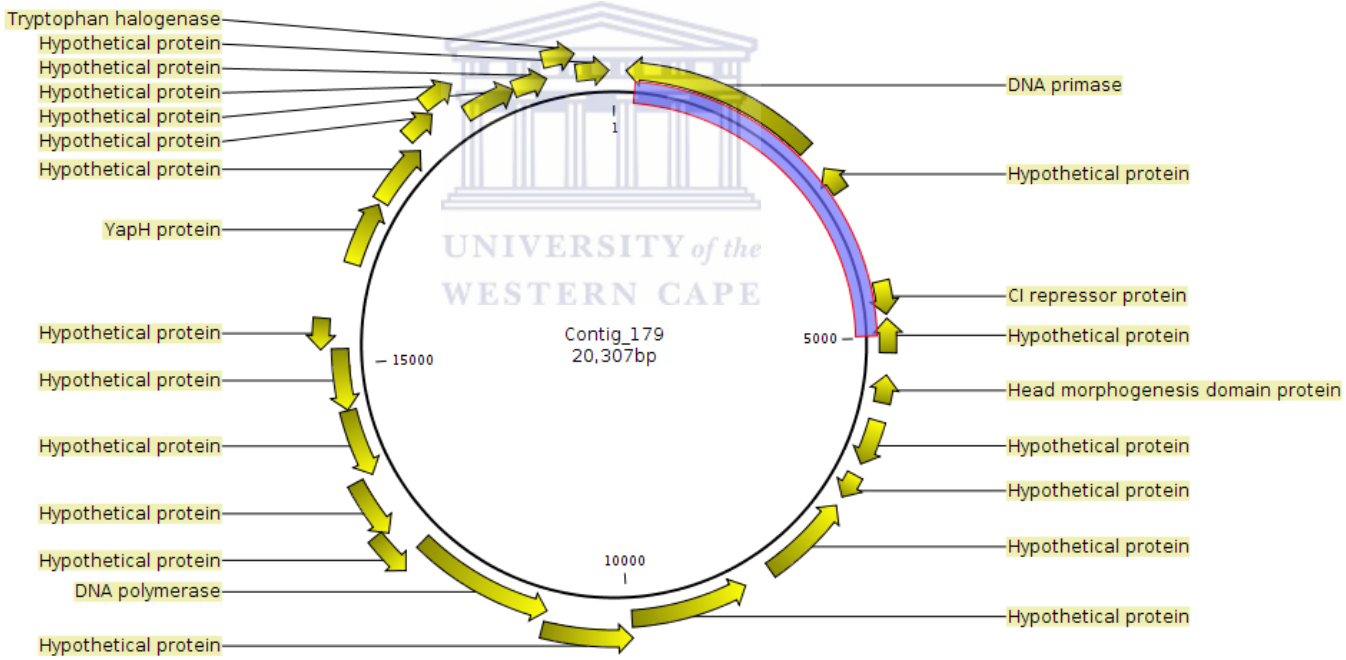

Figure 4.13: A graphical representation of contig 179 showing the open reading frames. The linear genome was computationally circularized for compact representation. The fragment highlighted in blue represents the 4.8kb fragment amplified from the metavirome.

Specific primers were also designed to amplify a 300 bp fragment of the genome of the novel *Salinivibrio*-infecting SMHB1 phage described in chapter three. The lysogenic nature of the phage provided an ideal system to test the FISH protocol on environmental samples. The probe was designed to a region of the terminase gene and designated MHB (Table 4.1).

The DS5 DNA fragment used as probe and target was amplified from metaviromic DNA and confirmed via sequencing. Smaller fragments were generated through restriction digestion (Figure 4.14) and 3' end labelled with cy3-dCTP using Terminal Deoxynucleotidyl Transferase, Recombinant (rTdT).

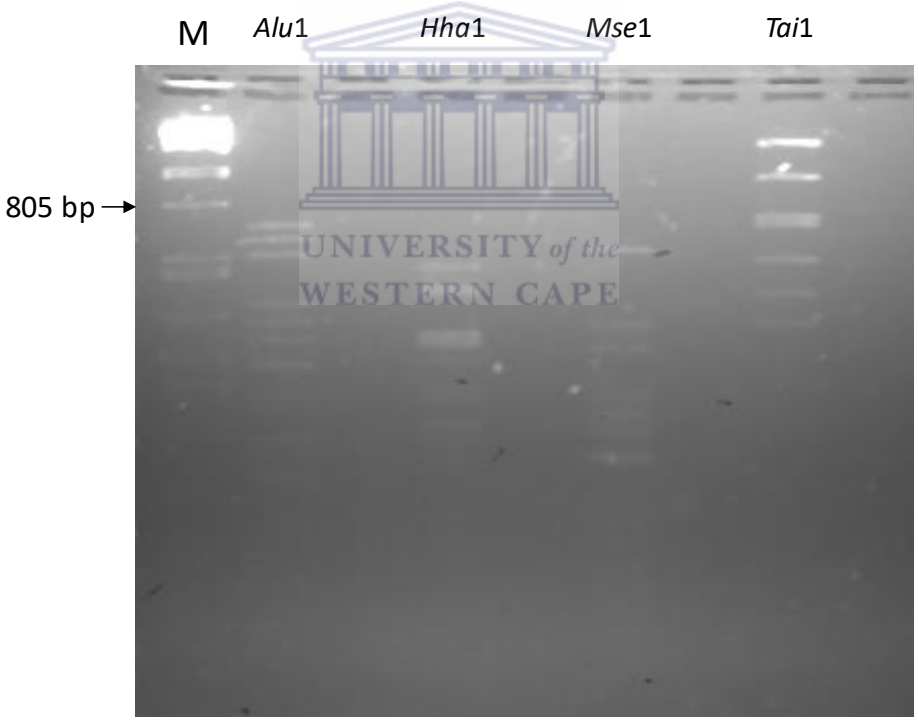

Figure 4.14: Restriction digest of the DS5 fragment using enzymes to produce FISH probes of random lengths having a mix of fragments with blunt ends (*Alu*I, *Mse*I) and 3' overhangs (*Hha*I, *Tai*I). Lambda DNA digested with *Pst*I was used as molecular marker.

4.3.2.2 Control hybridization experiment using NaCl based buffer

Standard FISH using a NaCl based hybridization buffer was used as the starting point for optimizations with a formamide concentration of 35%. Initial experiments involved probing *E. coli* BL21 cells containing the 4.8kbp fragment cloned in pJET (4.8kb/pJET/BL21) with 3' labelled, random length DS5 probes. The experiment showed relatively low fluorescence signal compared to the negative control under microscopic inspection (Figure 4.15).

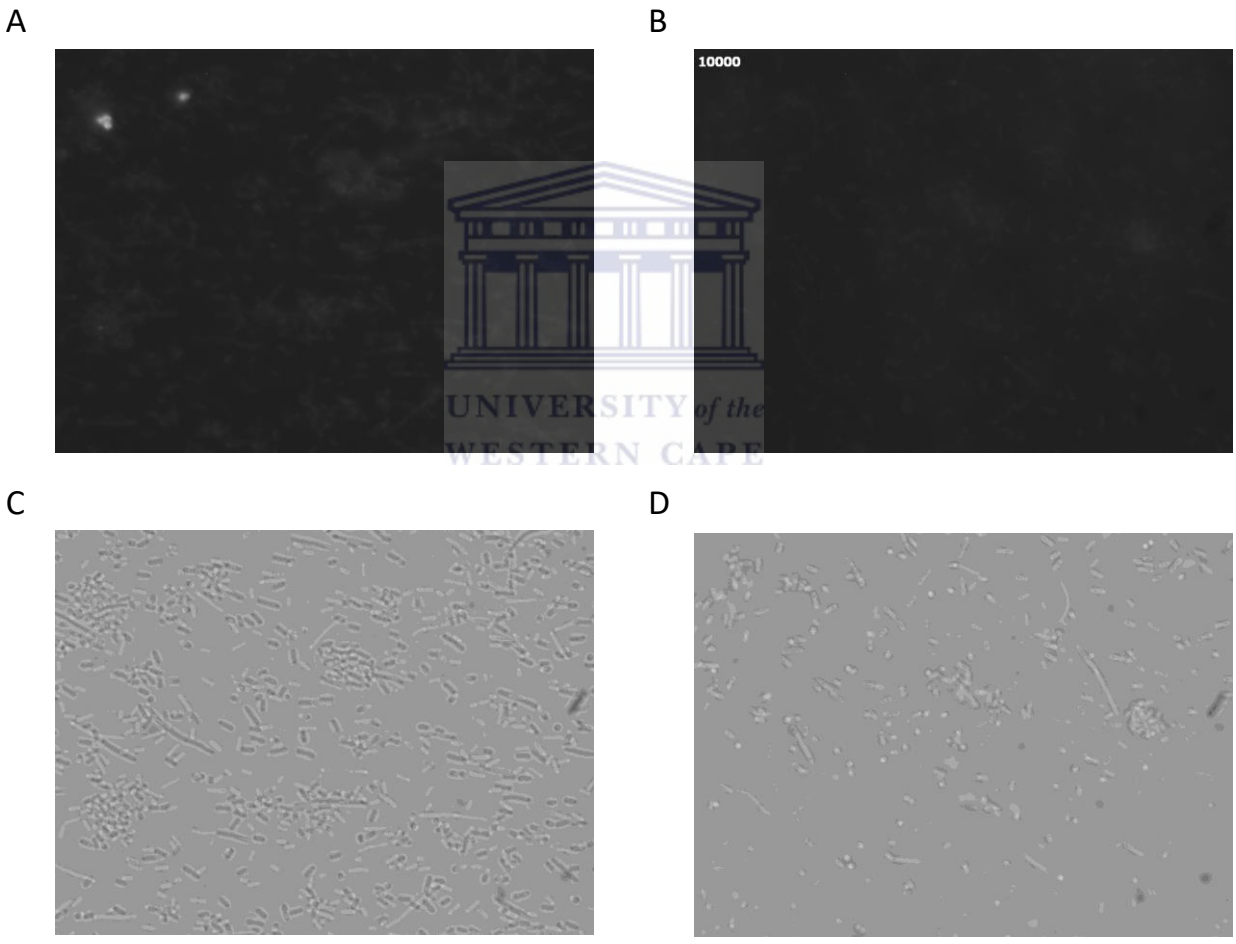

Figure 4.15: Micrographs showing 4.8kb/pJET/BL21 *E. coli* cells, with multiple copies of the target DNA fragment, after hybridization experiments. A. Cells with 3' labelled, random length probes under fluorescent light. B. Cells with unlabelled probe added under fluorescent light. C & D. Both experiments under bright field.

Similar experiments were carried out to improve the fluorescence intensity difference between positive and negative experiments. These included using a range of formamide concentrations between 35 – 55%, increasing the hybridization temperature by 1ºC every 4 hours over a range between 44ºC and 47ºC, and increasing the final probe concentration from 5 pg/µl to 5 ng/µl. Varying these parameters did not significantly improve the signal to noise ratio, based on microscopic observation.

4.3.2.3 PCR incorporation of CY3-dCTP

Incorporating the fluorescent dye into the full length 4.8kb DNA fragment *via* PCR before restriction digest was attempted to increase the fluorescent signal generated by each probe molecule. While 3' end labelling effectively meant one CY3 tag per probe strand, PCR labelling should result in the incorporation of multiple labelled dCTPs along the entire probe length, significantly increasing the fluorescence signal generated per strand.

DreamTaq and ExTaq DNA polymerases were tested to see which was more efficient at incorporating labelled dNTP. Phusion and phi29 DNA polymerases were also tested due to their ability to amplify difficult templates and yield higher amplicon concentrations.

However, probe labelling using DNA polymerases proved inefficient (Figure 4.16). Given the 3:2 ratio of dCTP to Cy3-dCTP in the dNTP mix, DreamTaq (lane 4), ExTaq (lane 8) and Phusion (not shown) polymerases seemed to preferentially incorporate unlabelled dCTP into amplicons during PCR. This was not unexpected but checking the amount of CY3 incorporated into each amplicon using a fluorometer gave a value of ≤ 0.2 pmol/µl. This was quite low compared to published experiments that required at least 7 pmol/µl for successful hybridization (Lieu *et al*., 2005; Redon *et al*., 2009). This suggested that the majority of the labelled dNTP remained unincorporated.

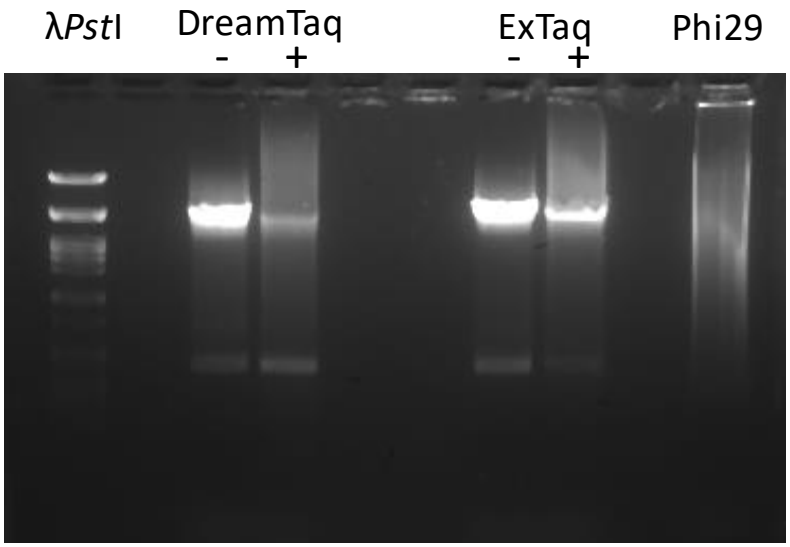

Figure 4.16: Cy3 incorporation into FISH probe using various DNA polymerases. Lanes with the + sign indicate reactions where a 3:2 ratio of dCTP to Cy3-dCTP was in the dNTP mix. The – sign indicate the absence of CY3-dCTP in mix.

While Phi29 DNA polymerase has been shown to be useful for incorporating labelled dNTPs into DNA (Smolina *et al*., 2005; Korlach *et al*., 2008), properties such as concatenation and supercoiling make the use of probes generated by rolling circle amplification in hybridization experiments more complicated. Such amplicons may not readily transverse the cell membrane during hybridization and stringency washing and may contribute to background fluorescence.

The DIG probe labelling system was therefore finally used for two reasons. First, the difficulty in generating sufficiently labelled DNA fragments of varying lengths using the digested DS5 probe as described above. Second, the DIG system has been optimized such that a DIG molecule is incorporated every 3-4 base pairs. This combines probe amplification and labelling into a single PCR reaction. When coupled with a signal amplification step, a single probe should generate high enough fluorescent signal intensity.

Other optimizations were introduced based on the choice of using the DIG labelling system. The hybridization and annealing conditions for each fragment of DNA is dependent on its length and GC content (Allers *et al*., 2013; Pan *et al*., 2014). Therefore, given that a single DIG

labelled probe should produce high fluorescence intensity, it was not necessary to attempt to find an optimum condition where a significant amount of probe molecules of random lengths of the 4.8kb fragment would hybridize sufficiently to the target to produce optimal fluorescence. It was therefore decided to use single specific probes. The probes were restricted to 300 bp as previously demonstrated to be the optimum length (Allers *et al*., 2013) and a melting curve analysis used to minimize the number of variables during DNA-DNA hybridization. The hybridization buffer was also changed from a NaCl to a SSC based composition as it has been optimized to incorporate other reagents that enhance the efficiency of probe hybridization (Allers *et al*., 2013). Such reagents include dextran sulphate, a high molecular weight polysaccharide, which minimizes background fluorescence and reduces non-specific binding by filling up the intracellular space (Moraru *et al*., 2010). Other compounds such as commercial blocking agents, yeast RNA and sheared salmon sperm DNA, play similar roles. These compounds were all incorporated into the hybridization buffer.

4.3.2.4 DIG labelling

Probes DS and MHB were therefore designed and successfully amplified, then DIG labelled. DIG incorporation modifies the electrophoretic mobility of DNA, hence, the successful labelling was assessed using agarose gel electrophoresis (Figure 4.17). As can be seen by comparing labelled and unlabelled probe, the labelled probe was positioned slightly higher on a gel.

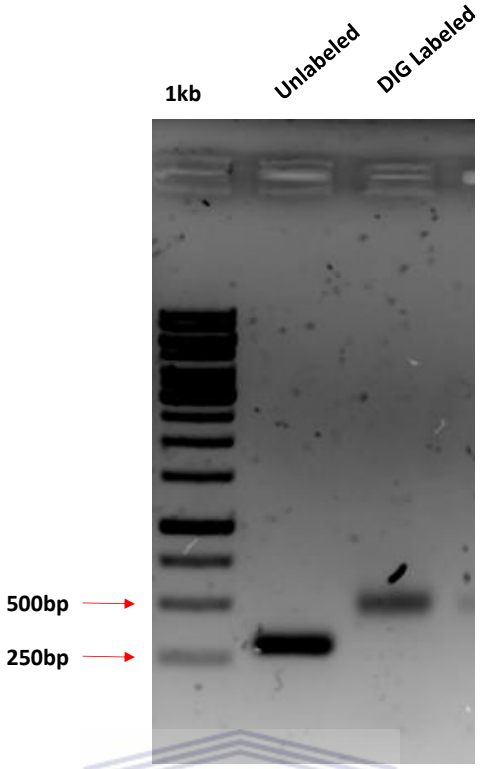

Figure 4.17: Visualization of DIG incorporation using a 2% agarose gel. DIG incorporated probe was positioned slightly higher than the unlabelled control due to reduced electrophoretic mobility, an indication of successful DIG incorporation.

4.3.2.5 Determination of optimum hybridization conditions using SSC based buffer

Using a melting curve analysis (Figure 4.18), the melting temperature (T_m) for the DS probe target hybrid in the hybridization buffer was determined to be 68^{o}C at 35% formamide and total Na^+ concentration of 1718 mM (Figure 4.18). The T_m was obtained when the probe was in wash buffer II. The hybridization temperature (T_{hyb}) was calculated as T_m -25^{o}C (Moraru *et al*., 2010), therefore 43^{o}C was used for hybridization experiments.

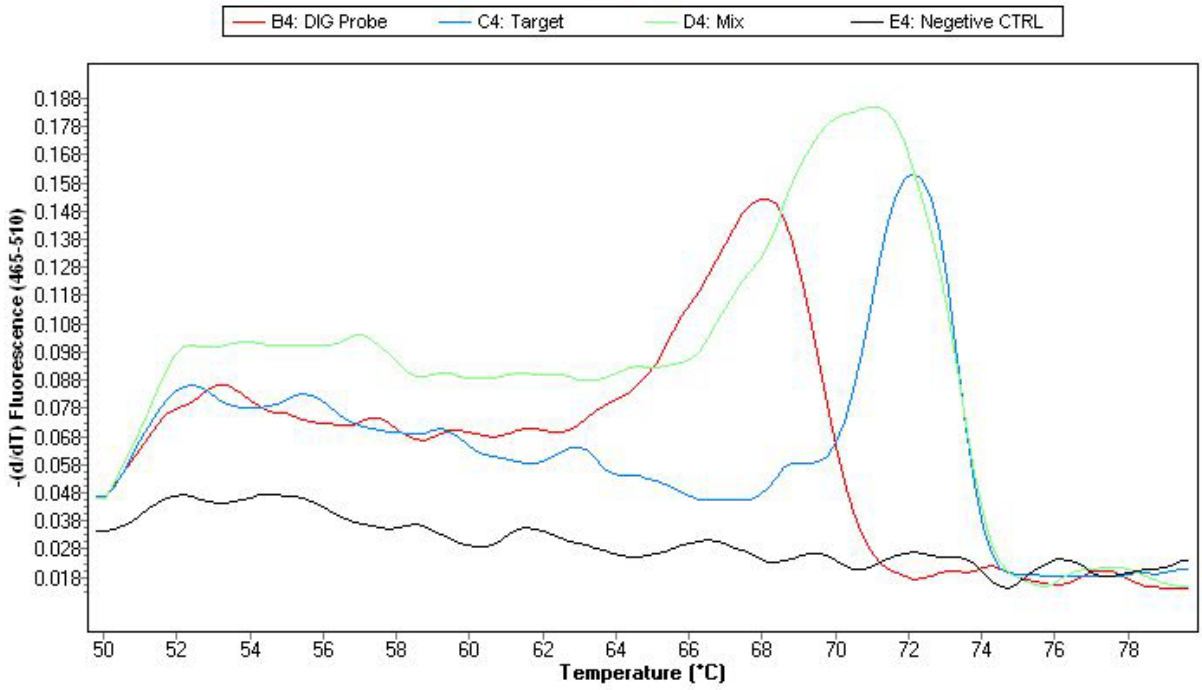

Figure 4.18: Melting curve analysis of DS probe-target hybridization in the hybridization buffer using fixed parameters. The reaction included a mixture of unlabelled and DIG labelled probe (1:1) (green). The unlabelled probe served as target. Control reactions included DIG labelled probe only (red), unlabelled probe only (blue), and a no DNA tube (black).

The calculated hybridization temperature differed significantly from the 46ºC recommended for standard FISH experiments. This is a critical factor in hybridization as a 1ºC change in temperature may be the difference between probe molecules remaining bound to the target or becoming unbound or loosely bounded and easily removed during stringency washes. (Pernthaler *et al*., 2002; Fuchs *et al*., 2010).

It is also of interest to note that similar hybridization temperature values were obtained for both the DS and MHB probes despite the significant difference in GC content (53% compared to 41%). The impact of the difference in GC content could have be negated by the presence of formamide and SDS in the buffer (Rose *et al*., 2002) .

4.3.2.6 Control hybridization experiment using SSC based buffer

The optimized hybridization conditions were tested using the DS probe and DS/pJET/BL21 system and the CsaB/pJET/BL21 system as the negative control. Initial hybridization experiments were characterized by low signal to noise ratio when viewed using fluorescent microscopy (Figure 4.19).

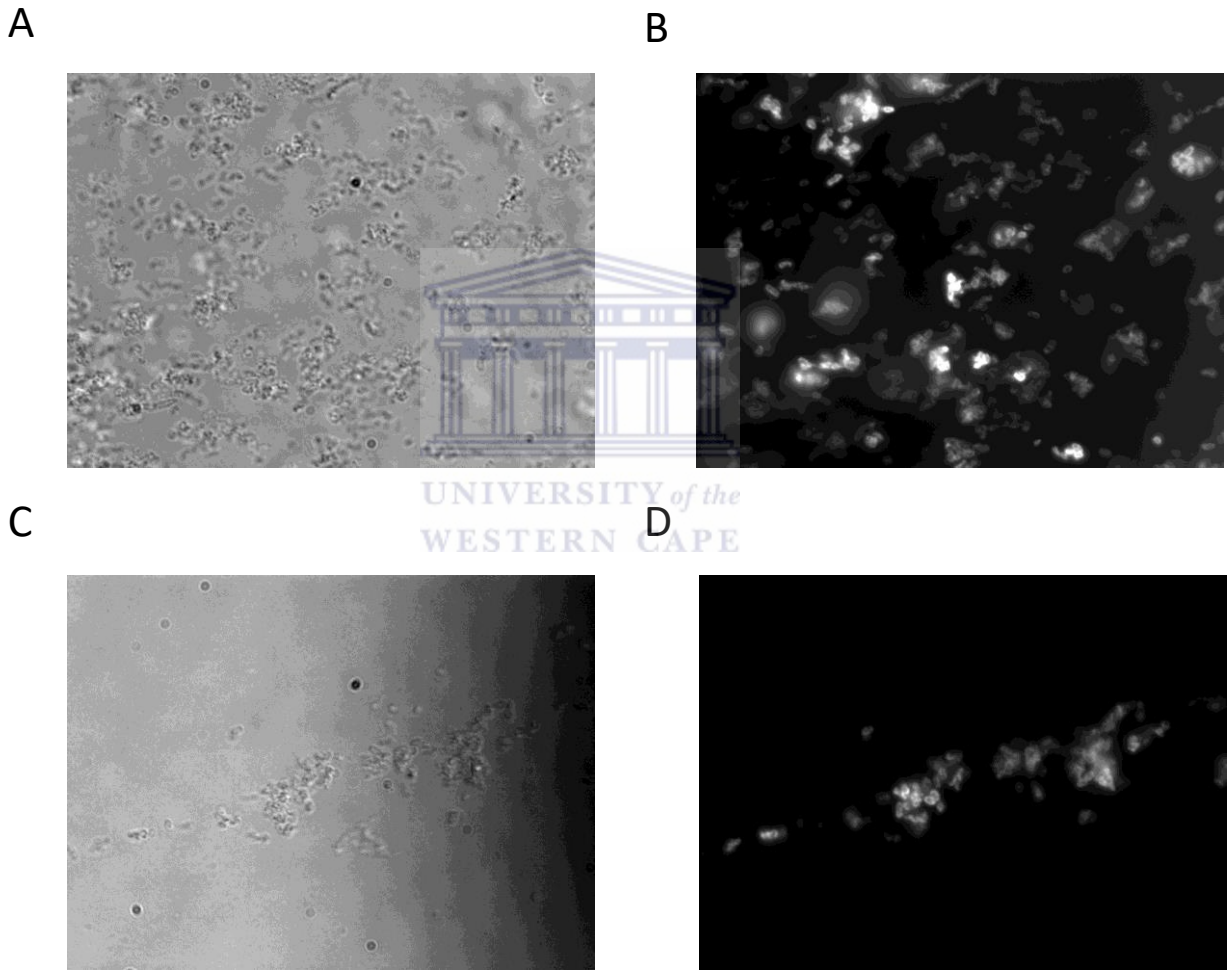

Figure 4.19: Micrographs showing *E. coli* cells, with multiple copies of the target DNA fragment, after hybridization experiments under conditions determined using a melting curve analysis. A. DS/pJET/BL21 cells with DS DIG labelled probe added under bright field. B. DS/pJET/BL21 cells with DS DIG labelled probe added under fluorescent light. C & D. CsaB/pJET/BL21 negative control with DS DIG labelled probe.

The samples were then analyzed using flow cytometry, where a difference in fluorescence signal was observed (Figure 4.20). As shown in Figure 4.20A, approximately 84% of target cells (DS/pJET/BL21) showed increased fluorescence compared to 77% of cells without the target (*CsaB*/pJET/BL21). Such a difference would have been difficult to observe under a fluorescent microscope. Therefore, from this point onwards, samples were analyzed using only flow cytometry.

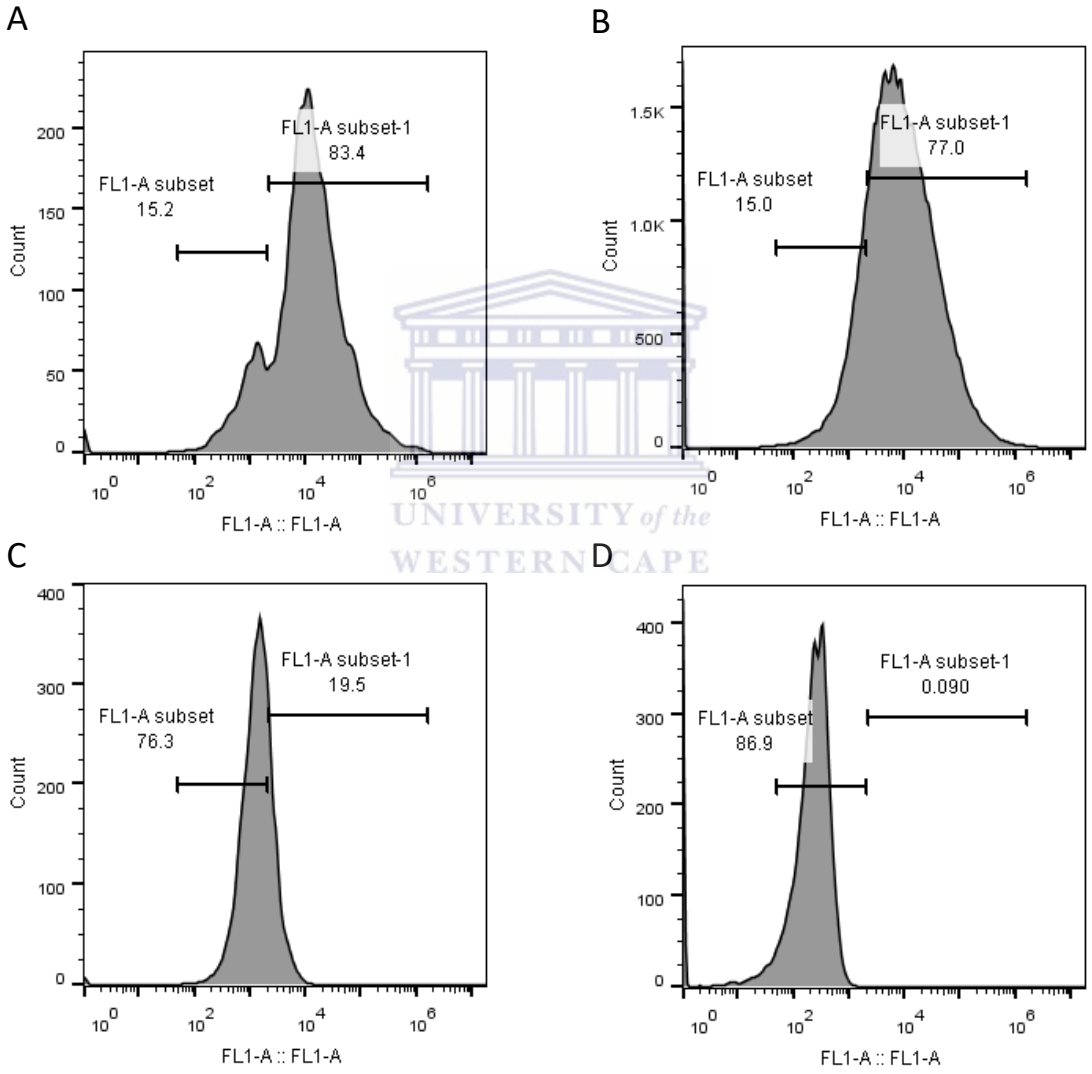

Figure 4.20: Histograms showing the fluorescence intensity of FACSorted *E. coli* experiments. The number of cells is represented on the y- axis, while the x-axis represents fluorescence within the FL1 (488nm) laser detector of the flow cytometer. A. DS DIG labelled probe against complementary DS/pJET/BL21 target B. DS DIG labelled against non-complementary CsaB/pJET/BL21 target C. BL21 cells treated as in A and B but without probe D. Untreated FACSorted BL21 cells.

Experiments were also conducted to measure the impact of background fluorescence from cells. Two controls were included to measure fluorescence from untreated *E. coli* BL21 cells vs cells which were exposed to the hybridization process but without probe. As shown in Figure 4.20C, 19% of the treated *E. coli* cells had some fluorescent signal while less than 1% of untreated cells had any (Figure 4.20D). This showed that the effect of inherent fluorescence in *E. coli* cells was negligible. However, the relatively high fluorescence observed for the non-probe complementary *CsaB*/pJET/BL21 cells (77%, Figure 4.20B) showed that the hybridization process required more optimization to avoid increasing the background fluorescent signal.

A number of optimizations were introduced in subsequent experiments to further improve the resolution of the fluorescence signal. These included i) the amplification of fresh probe each time to avoid the potential effect of a freeze-thaw on attached DIG molecules; ii) the direct use of amplified probe for hybridization without passing through a purification kit as suggested by the probe synthesis kit manufacturers; and iii) the use of freshly prepared solutions wherever possible. Others included the use of another tyramide signal amplification buffer (Moraru *et al*., 2010) instead of the buffer provided with labelled tyramide, and the increased incubation time for signal amplification from the recommended 10 minutes to 45 minutes.

These optimizations resulted in an improved resolution of the FISH signal. As shown in Figure 4.21 below, approximately 87% *of E. coli* cells (DS/pJET/BL21) containing the cloned target showed at least a ten-fold increase in fluorescence compared to cells with a non-complementary target (33.6%). Given that these samples were treated the same, the obvious difference in fluorescence is an indication of successful probe – target hybridization. The smaller peak (red arrow, Figure 4.21A) most likely represented un- or poorly labelled cells. This is likely an indication that the hybridization washes were stringent enough or could represent cells which have lost the plasmid.

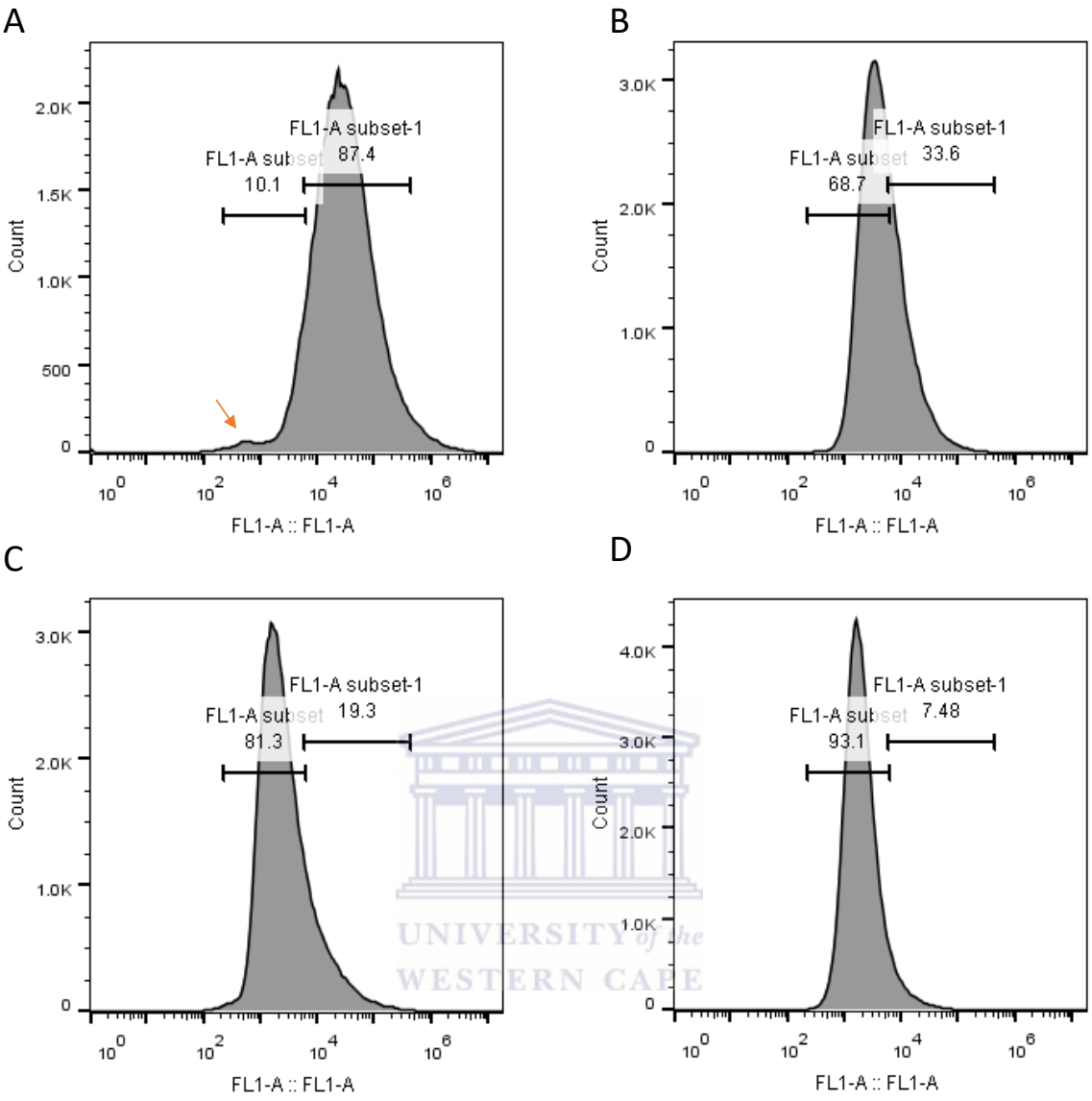

Figure 4.21: Histograms showing the fluorescence intensity of FACSorted control *E. coli* BL21 experiments. A. DS DIG labelled probe against complementary DS/pJET target B. DS DIG labelled against non-complementary CsaB/pJET target C. Unlabelled DS probe against DS/pJET target D. Unlabelled DS probe against non-complementary CsaB/pJET target.

A dot plot of cell size (forward scatter, FSC) against fluorescence was constructed for the events above to identify cells with a distinct shift in fluorescence (Figure 4.22). The gate was constructed based on the unlabelled population recorded for non-complementary probe targets (Figure 4.22D).

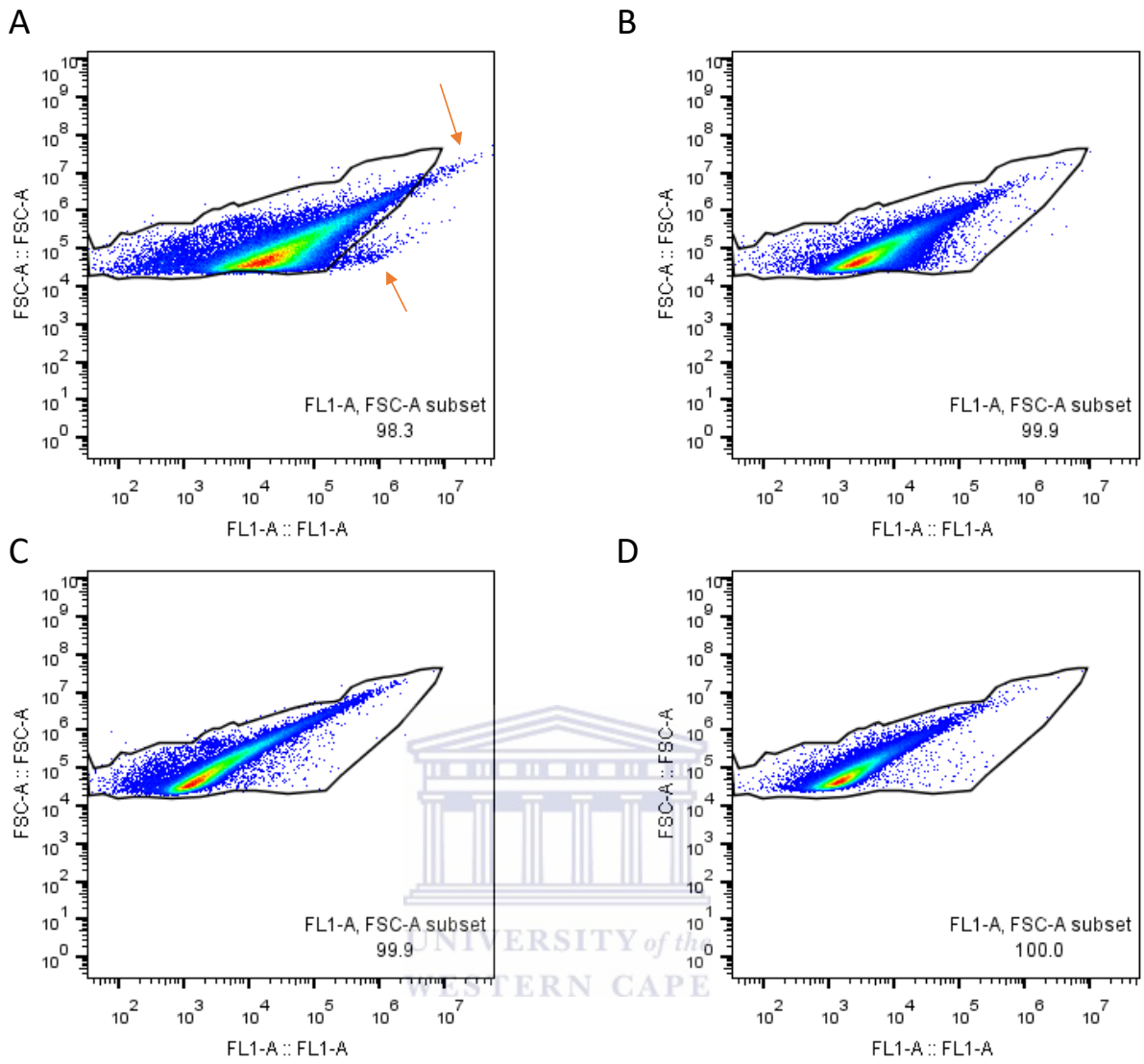

Figure 4.22: Dot plots showing the fluorescence intensity of FACSorted *E. coli* BL21 experiments. A. DS DIG labelled probe against complementary DS/pJET target B. DS DIG labelled against non-complementary CsaB/pJET target C. Unlabelled DS probe against DS/pJET target D. Unlabelled DS probe against non-complementary CsaB/pJET target.

The clear shift in the experimental population versus the control populations shows that the hybridization and stringency wash conditions had now been tuned to a point where these populations could be distinguished from one another. Two distinct populations were identified (red arrows, Figure 4.22A) with the cells in one being at least one order of magnitude larger than the other. A possible explanation is that the cells in the 'bigger' population were close to the end of cell division and therefore could be two daughter cells that were harvested before

separation. The fact that the difference in fluorescence intensity between the two populations is also approximately one order of magnitude higher supports this as two joined daughter cells would imply twice the copy number of the plasmids, hence a higher number of FISH targets per 'event'.

A few false positives (< 20 events) were observed for the non-complementary *CsaB*/pJET/BL21 cells (Figure 4.22B) and DS/pJET/BL21 target cells to which unlabelled DS probe was added (Figure 4.22C). This could be attributed to minimal residual labelled tyramide or non- specific hybridization.

These experiments provided the proof of concept of the possibility of being able to use FISH to discriminate between target and non-target cells in a controlled setup. However, these experiments were conducted with an abundance of targets per cell (400-700 plasmid copies), a condition unlikely to exist in the environment which was the focus of the study. It was therefore necessary to evaluate the specificity of the optimized protocol to reflect what is obtained in the environment (1-4 targets per cell). It was also critical to determine the detection limit of the fluorescent signal with respect to the number of target cells within the environmental bacterial population. This would inform on the resolution power of the optimized protocol.

4.3.2.7 Establishing hybridization specificity and signal detection limit

To test for hybridization specificity, the protocol was applied to differentiate between the two *Salinivibrio* species, SS2 and SS3, isolated from the Swakop saline site (Chapter two). Strain SS3 has been shown to contain a single copy of the genome of the lysogenic phage SMHB1 integrated within its genome. Host range testing and PCR analysis confirmed that the phage did not infect strain SS2 (Section 3.3.2). This therefore provided an ideal environmental system

where closely related bacterial strains could be differentiated on the basis of a single copy target.

The strains were probed using the SMHB1 phage specific probe, MHB. The optimum hybridization conditions were determined in a similar manner to the DS probe and found to be identical (Section 4.3.2.5). The same hybridization conditions were therefore applied. The DS probe was used as the non- specific control probe. The strains were hybridized individually and also as a 1:1 mix.

An analysis of the fluorescence profile of strain SS3 when probed with the DIG labelled, phage specific, MHB probe revealed an increase in fluorescence, with 79% for the phage host strain SS3 showing a shift (Figure 4.23A). A comparison of these cells with the mixed population probed with unlabelled MHB (Figure 4.23E) showed that a significant amount of the MHB probe did bind specifically within the SS3 population (79% against 23%). This relatively low percentage (23%) also showed that while there was some background fluorescence, the MHB probe hybridized sufficiently enough to distinguish between labelled and unlabelled cells.

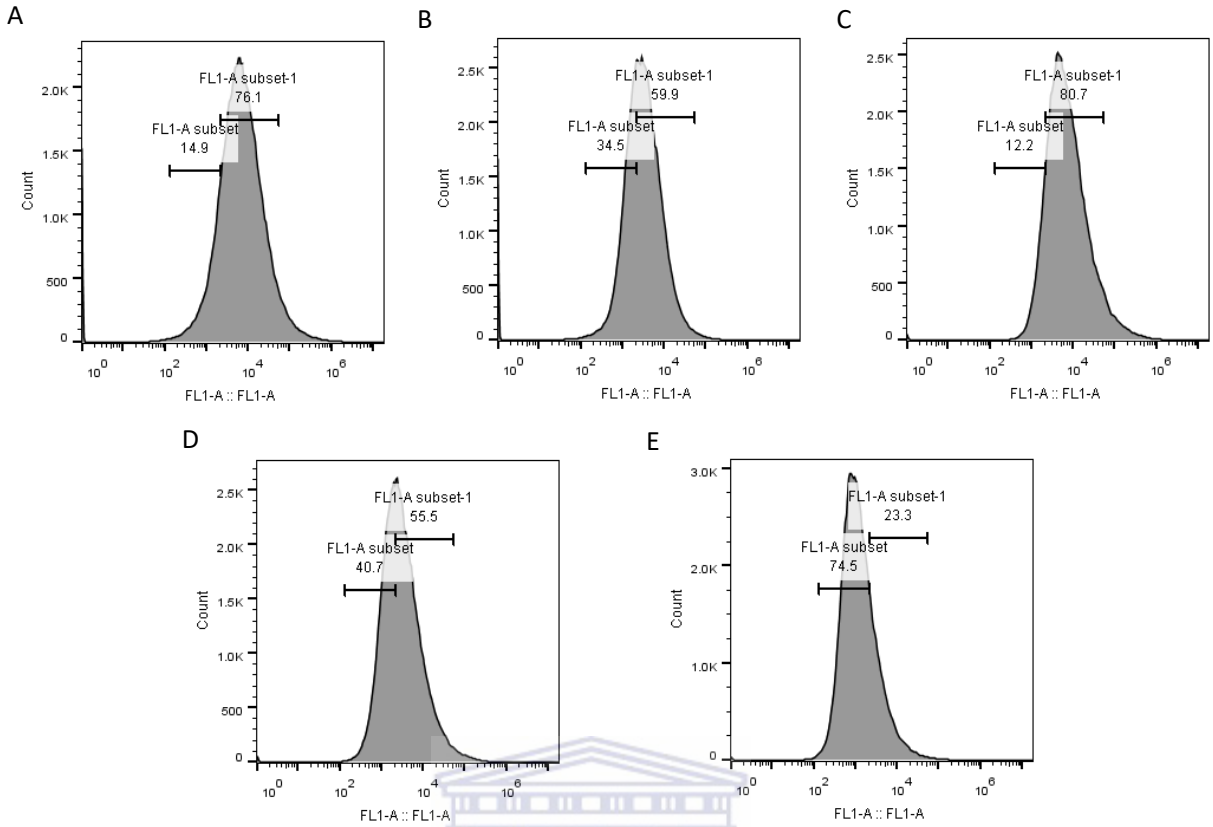

Figure 4.23: Histograms showing the fluorescence intensity of mixed *Salinivibrio* strains. A. Strain SS3 with the phage specific, DIG labelled MHB probe B. Strain SS2 with the phage specific, DIG labelled MHB probe C. Mixed strains with the phage specific, DIG labelled MHB probe D. Mixed strains with the non- specific, DIG labelled DS probe E. Mixed strains with the phage specific, unlabelled MHB probe.

Strain SS2 also showed some increase in fluorescence (60%, Figure 4.23B), however, the signal intensity was similar to the mixed population probed with the DIG labelled but non-specific DS probe (56%, Figure 4.23D). The similarity observed suggests that some of the fluorescent signal was due to non- specific binding. The effect of non-specific binding could also explain why although the highest fluorescent signal was observed in the mixed population probed with DIG labelled MHB probe (81%, Figure 4.23C), a single peak was observed showing that the strains remained indistinguishable.

The experiment above was also analyzed using the dot plot function to determine the amount of fluorescence per event and to attempt finding a population showing a distinct shift. Such a shift would represent target cells to which the most number of probe molecules have hybridized.

A comparison of the dot plots (Figure 4.24) showed a population just outside the gate (red arrows), which is what one would expect to represent the population that has hybridized to the DIG-labelled MHB probe. However, a similar population was found for the SS2 cells which do not have the target (Figure 4.24B), which also suggests non-specific hybridization.

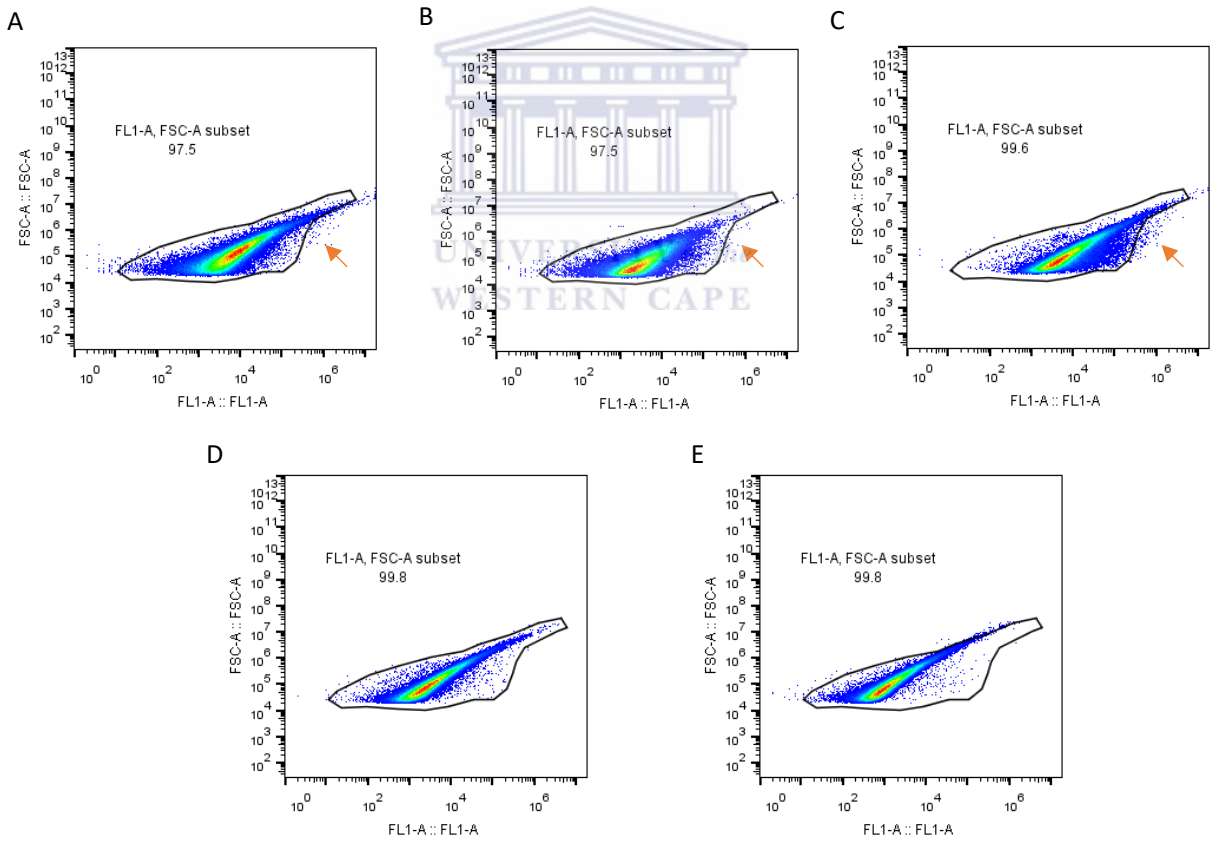

Figure 4.24: Dot plots showing the fluorescence intensity of mixed *Salinivibrio* strains. A. Strain SS3 with the phage specific, DIG labelled MHB probe B. Strain SS2 with the phage specific, DIG labelled MHB probe C. Mixed strains with the phage specific, DIG labelled MHB probe D. Mixed strains with the non- specific, DIG labelled DS probe E. Mixed strains with the phage specific, unlabelled MHB probe.

The absence of that population in the controls which include cells probed with the non-complementary DV probe (Figure 4.24D) and the unlabelled but complementary MHB probe (Figure 4.24E) showed that the fluorescence observed in that population is unlikely to be due to residual labelled tyramide.

To further understand the effect of non-specific hybridization and the target detection level within a community, these experiments were repeated with a mock bacteria mix. To simulate an environmental bacterial fraction, isolates from the salt playa were grown in a range of media (Chapter two) and the ratio of OD to viable cells determined using the standard plate count method. The mock mix therefore contained the following strains: *Halomonas caseinilytica, Halomonas eurihalina, Idiomarina loihiensis, Marinobacter xestospongiae, Virgibacillus salarius* and SS2 (*Salinivibrio sp.*). The experiment was first carried out using the control DS/pJET/BL21 *E. coli* cells which should be easier to detect because of the high copy number of the target in the cells. The cells were added to the mock mix cells in ratios 1:5 and 1:100. The experiment was repeated using the phage host, SS3, added in similar ratios.

For the *E. coli* experiment, a pattern similar to the earlier experiment where two distinct populations were observed, was observed in the 1:5 cell mix (Figure 4.25A & B). The pattern appears to correlate with the number of target cells present per sample with the sample containing pure DS/pJET/BL21 *E. coli* cells having a higher number of fluorescent events outside the gate followed by the 1:5 mix and the 1:100 population. Interestingly, just two events within the negative control (Figure 4.25E) were outside the gate. This showed that it is possible to select for the exact population containing the sequence of interest within a mixed population, if they are well represented. However, where the target is present in low abundance it will be difficult to differentiate the hybridized population from the background fluorescence of the mixed community.

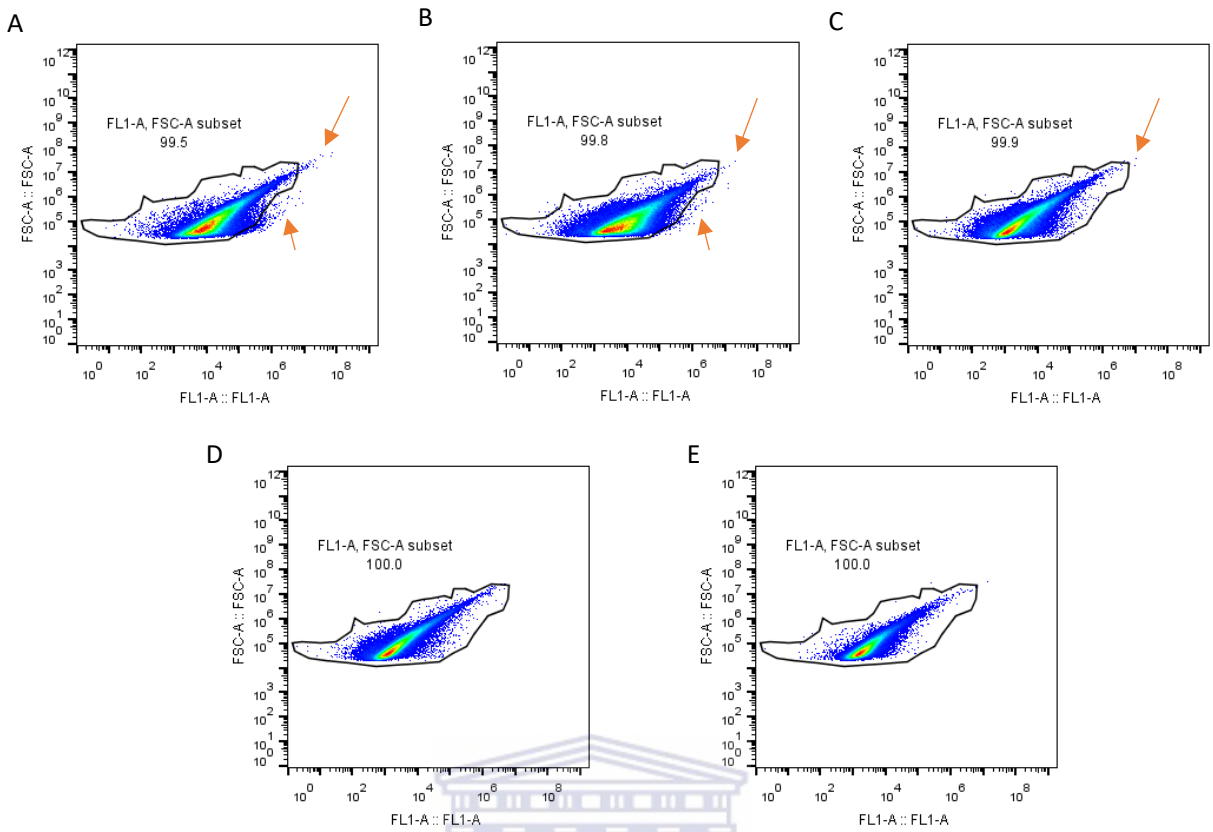

Figure 4.25: Dot plots showing the fluorescence intensity of *E. coli* clones within a mix of bacterial cells after hybridization using DIG labelled DS probe. A. DS/pJET/BL21 cells B. DS/pJET/BL21 cells in a 1:5 mix with other bacterial isolates C. DS/pJET/BL21 cells in a 1:100 mix with other bacterial isolates D. Mock bacterial mix without DS/pJET/BL21 cells E. Non-complementary CsaB/pJET/BL21 *E. coli* cells.

Comparing the histograms of the experiments showed a significant reduction in fluorescence from the 1:5 mix (51%, Figure 4.26B) to the 1:100 (14%, Figure 4.26C). This is most likely a reflection of the number of target cells. The fluorescence intensity of the bacterial mix without target DS/pJET/BL21 cells is represented in Figure 4.26D and given that the fluorescent signal should include both the residual intrinsic fluorescence from the cells after treatment and any non- specific hybridization from the addition of probe, a value of 12% shows that the effect of background fluorescence was minimal. This is further supported by the low percentage (8.5%) of cells fluorescing when the same conditions where applied to the CsaB/pJET/BL21 cells which represent a non-specific, but homogenous population. It is clear that the majority of the

fluorescent signal comes from the DS/pJET/BL21 *E. coli* cells as reflected when comparing the samples where it is present (Figure 4.26A, B & C) against where it is absent (Figure 4.26D & E).

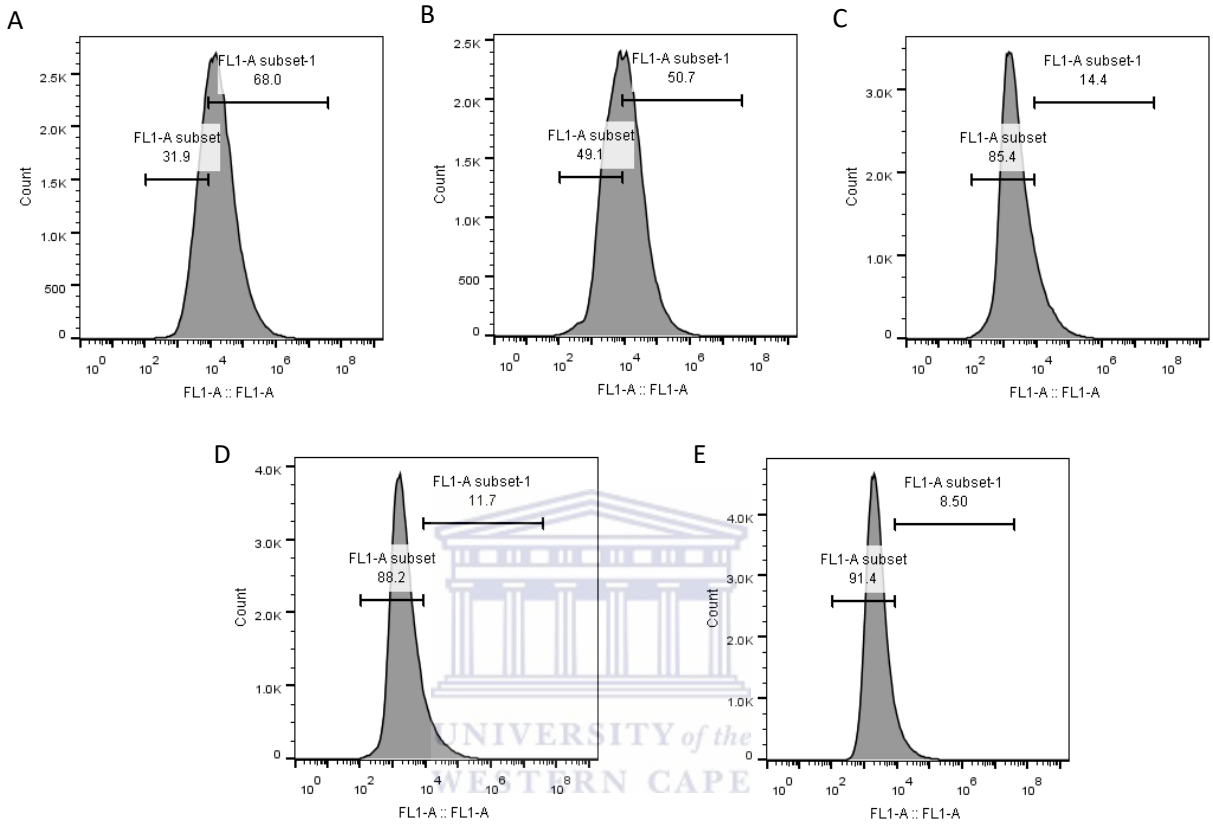

Figure 4.26: Histograms showing the fluorescence intensity of control *E. coli* cells within a mix of bacterial after hybridization using DIG labelled DS probe. A. DS/pJET/BL21 cells B. DS/pJET/BL21 cells in a 1:5 mix with other bacterial isolates C. DS/pJET/BL21 cells in a 1:100 mix with other bacterial isolates D. Mock bacterial mix without DS/pJET/BL21 cells E. Non-complementary CsaB/pJET/BL21 cells.

The relative similarity between the 1:100 mix (14%, Figure 4.26C) and the bacterial mix where the target is absent (12%, Figure 4.26D) shows the detection limit of the current protocol. This was further confirmed when the experiment was repeated using SS3 cells as target. The trend observed was different from the *E. coli* clones (Figure 4.27). The fluorescence signal was reduced which was expected because of the difference in the number of probe targets per cell, however, there was no difference between the 1:5 (26%) and the 1:100 mixes (27%). A

difference in fluorescence intensity was observed when comparing samples containing the target (Figure 4.27A, B & C) with samples where the target is absent (Figure 4.27D & E), but the difference is not sufficient enough to confidently identify the target population.

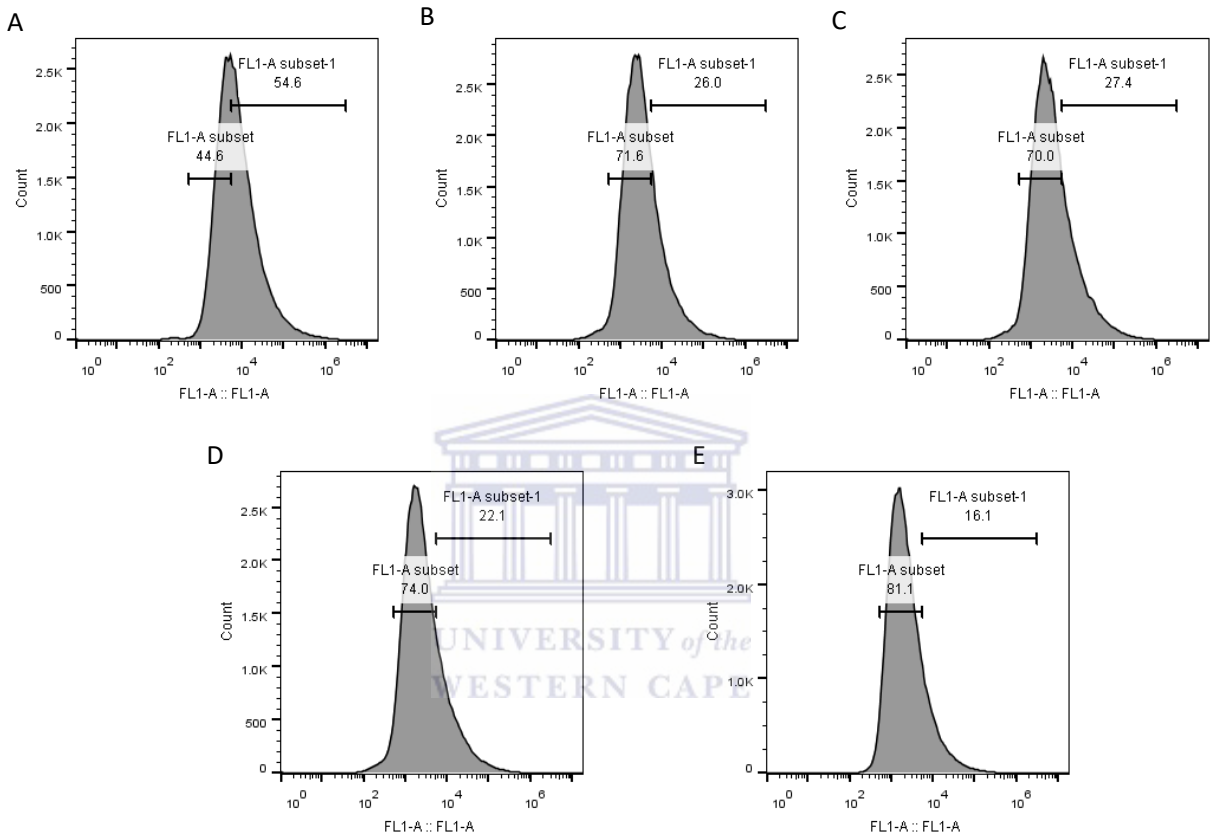

Figure 4.27: Histograms showing the fluorescence intensity of SS3 within a mix of bacterial cells after hybridization using DIG labelled, phage specific MHB probe. A. SS3 cells B. SS3 cells in a 1:5 mix with other bacterial isolates C. SS3 cells in a 1:100 mix with other bacterial isolates D. Bacterial mix without SS3 cells E. SS2 cells.

4.3.3 IMPLICATIONS OF THE METHODOLOGY FOR ENVIRONMENTAL SAMPLES

The findings in this study raise the question of how best the method may be applied to a typical environmental mixed community. First, it poses the question of how to go about selecting the best target. On the one hand, the study has shown that target copy number significantly impacts the resolution of FISH signal, however, given that the most abundant bacteria tend to have lytic rather than lysogenic phages (Thingstad, 2000; Clokie *et al*., 2011), they represent unlikely FISH targets. On the other hand, while low abundant bacterial species usually harbor lysogenic phages (Thingstad, 2000; Clokie *et al*., 2011), there seems to be a population-dependent threshold for FISH to be successfully applied. A useful strategy might be to try to infer the relative abundance of the host from metaviromic sequence data. Given that phages outnumber bacterial hosts by 10:1 (Chibani-chennoufi *et al*., 2004; Stern *et al*., 2012), a moderately covered phage genome within the dataset might suggest a similar abundance of the host in that environment.

Second, the heterogeneous nature of an environmental bacterial community suggests that some of the species might not be amenable to the treatments (for example fixation permeabilization, and peroxidase inactivation) employed in this study, and may be lost. It is therefore necessary to test a range of treatments that are effective while keeping cell loss to the minimum.

4.4 CONCLUSION

Fluorescent labelling has been applied to viral studies as stated earlier, but this study has demonstrated the difficulty that might be faced when applying such techniques to environmental samples. While the detection limit needs to be improved, to the best of our knowledge, this is the first study to show that phage bacterial host can be detected in a mixed population using FISH.

The detection limit can be improved by either increasing the fluorescent signal per cell or further reducing the background fluorescence. The use of multiple probes could be explored although each probe combination will require additional optimization to avoid non-specific binding. The use of probes with fluorescent tags that have a signal quenching mechanism may also improve the signal to noise ratio (Silahtaroglu *et al*., 2004). Such probes will also reduce the number of downstream processing steps, preserve cell morphology and save time. Reducing background fluorescence might prove more difficult for environmental fractions as the biology of the bacterial species is not known. For example, increasing the concentration of exposure time to HCl used for inactivating endogenous peroxidases (known to fluoresce) may result in the loss of some novel species. Also, it was observed in this study that stringency washes made cells translucent, weakened integrity and resulted in significant loss of intact cells, however, reducing the wash time resulted in increased background fluorescence. Alternative washing methods that rely less on centrifugation might be required.

There are clearly a number of challenges but FISH offers a promising approach to isolating novel phage-host pairs from environmental samples.

5

GENERAL DISCUSSION AND CONSIDERATIONS

The study of phage-bacterial interaction is of significant interest to the scientific community for a number of reasons. Phages, being the most abundant biological entities, directly impact ecological processes such as biogeochemical cycling and bacterial turnover on a global scale (Thurber, 2009; Ghannad & Mohammadi, 2012). The interaction of phages with the host is now understood to influence evolutionary trends in bacterial genomes (Sleator *et al*., 2008; Simon & Daniel, 2011). In addition, the genetic material within phage heads codes for several enzymes like DNA ligase, DNA and RNA polymerases that are of significant interest in biotechnology (Borysowski *et al*., 2006; Drulis-Kawa *et al*., 2012; Moser *et al*., 2012). However, phage studies have been limited by the difficulty in isolating novel species, a direct consequence of the unculturability of most bacterial hosts (Vartoukian *et al*., 2010). There is therefore a need to develop new methods to better access the diversity and improve our knowledge of phage-host interactions. Therefore, the aim of this study was to develop such novel high throughput techniques which, in tandem with genomic data analysis, should lead to an improvement on the current traditional phage isolation strategy.

Bioinformatic analysis of the phage diversity of two Namibian sites revealed a number of interesting observations. Analysis of the copper site metavirome showed that 67% of the predicted viruses were tailed affirming previous studies of soil viral diversity of extreme environments such as the Sahara, Mojave and Namibian deserts which showed that the viral populations are dominated by tailed morphotypes belonging to the families *Myoviridae*, *Siphoviridae* and to a lesser extent, *Podoviridae* (Prestel *et al*., 2008; Prestel *et al*., 2012; Prestel *et al*., 2013). However, non-tailed and short tailed (podoviruses) phages are thought to dominate aquatic environments because environmental factors such as salinity, tidal movements and ocean current cause tail loss and virus disruption (Wommack & Colwell, 2000; Brum *et al*., 2013). A global survey showed that 51-92% of viruses in marine environments are

non-tailed viruses (Brum *et al*., 2013). This was in contrast with the observation of the Swakop saline site where 66% of the predicted viruses were tailed, suggesting that the aforementioned factors may not necessarily be the primary drivers of phage diversity in this environment. The diversity within this environment therefore warrants further exploration.

Given biases introduced during sample preparation, especially in the extraction of viral genetic material, it is important to question whether we have captured the true diversity of these environments. Recent studies have shown that RNA viruses are nearly as abundant as dsDNA phages, representing 15-77% of viruses in some coastal waters (Steward *et al*., 2013; Brum *et al*., 2013). The dominant species are thought to be picorna-like viruses which are so small they fall below the detection limits of currently used fluorescence-based counting methods (Steward *et al*., 2013). These studies suggest we might be missing a significant proportion of the viruses present in the environment. Therefore, conclusions on viral diversity drawn from culture dependent or independent experiments that do not take into account non-dsDNA viruses might not be very accurate. To address this, genetic material extractions need to be more representative. A study successfully used hydroxyapatite chromatography to fractionate and sequence dsDNA, ssDNA, dsRNA, and ssRNA genomes from known phages, however, it was observed that regardless of the source nucleic acid, taxonomic analysis suggested that the phages were most similar to dsDNA viruses, a consequence of the bias in metaviromic databases (Andrews-Pfannkoch *et al*., 2010). Despite the bias, this holistic approach is currently the best way to capture the true diversity of any environment. Future experiments, including a focus on RNA viruses, may finally start to reveal the extent to which RNA viruses are represented in the environment and allow us to properly assess their diversity and ecology.

This study explored a number of approaches to isolating novel phage-host pairs which involved the application or optimization of techniques such as viral tagging and reverse metaviromics. Viral tagging was investigated as an approach to isolate a fluorescently-labelled phage adsorbed to its host through flow cytometry (Mosier-Boss *et al*., 2003; Deng *et al*., 2012). One limitation with this approach is that adsorption does not always mean infection as explained in Chapter one, hence there is a risk of mismatching a phage-host pair. Phage adsorption is time dependent, infecting phages bind irreversibly to the host and non-infecting phages will eventually detach (Moldovan *et al*., 2007). It could therefore be argued that increasing the time between challenging the bacterial fraction with labelled phages and sorting phage-host pairs would reduce the chances of false positives. However, given that for most environmental phages, the latent period is unknown, in a case where such a period is very short, the host would be lysed and lost before sorting. The approach might therefore be more applicable for targeting temperate phages.

The second limitation with viral tagging is the efficiency of labelling and the effect of the labelling agent on the ability of the phage to attach successfully to the host. This study found labelling of the environmental phage fraction to be problematic (Section 4.3.1) possibly due to interference from free cysteine residues (biotinylation) on the phage protein coat (Koide *et al*., 2009), and dye (SYBR gold) leaks associated with increased pressure within the phage head as a result of intercalation with the genetic material (Ohno *et al*., 2012). Microscopic evaluation and flow cytometry data suggested that not all phages within the environmental fraction suffered from dye leaks and cysteine interference, however, the diversity of such a sample suggests that the likelihood of this re-occurring could be quite high. For these reasons, the use of nucleic acid stains on an environmental fraction is probably not ideal. On the other hand, biotinylation may be improved by inactivating reactive cysteine residues. In preparing phage display libraries, such inactivation is achieved by mutagenesis or chemical modification (Koide

et al., 2009), but the effect of such agents on environmental phage fractions needs to be evaluated. For signal improvement, quantum dot conjugated streptavidin could be used instead of FITC as it has been shown be more photo stable and produce a higher noise to signal ratio compared to other fluorophores (Edgar *et al*., 2006).

The second approach, reverse metaviromics, was more target specific and relied on sequence data analysis as the starting point for novel phage identification followed by an attempt to identify and isolate the host. Identification of the bacterial host was investigated using two methods.

First, *in situ* phage capture. This bypasses concerns about the labelling agent interfering with phage attachment as the genetic material is already in the bacteria cell. Experimental approaches such as microfluidic PCR (Tadmor *et al*., 2011), phageFISH (Allers *et al*., 2013), and single cell genome screening (Martínez-García *et al*., 2014; Labonté *et al*., 2015) are based on the *in situ* detection of phage genetic material via PCR and/or hybridization. This means that prior knowledge of the phage sequence is required for primer or probe design, information that is readily available from metaviromic sequence data. Another pre condition for successful phage detection is that the phage is either lysogenic or has recently infected the host and the cell fixed before the lytic cycle is completed. This study attempted to identify the host of a potentially novel phage identified in the metavirome sequence, referred to as Contig 179, showing some similarity to a prophage of the cyanobacterium *Dactylococcopsis salina* (Chapter four). Experiments using the control *E. coli* system showed that 87% of the cells had at least a 10-fold increase in fluorescent signal, showing significant hybridization of the phage specific probe to the target, but simulation experiments had a detection limit of 1:100. The implications of such a low detection limit and a number of ways to improve it, such as the use of multiple probes and probes with signal quenching or activating mechanisms, were suggested in Chapter four. In essence, the technique as described here is not yet refined enough to deliver

on the promise of a high-throughput technique for the isolation of phage-host pairs. A recent study took an interesting approach to *in situ* phage host detection which did not require prior knowledge of either (Martínez-García *et al*., 2014). The viral genomes were cloned and immobilized while the bacterial genomes were sheared, size selected and used as probes. The probes were then hybridized to the immobilized clones. The approach successfully identified viruses infecting *Nanohaloarchaeota*, but the broad size range of the probes (0.4 to 1.6kb) coupled with the seemingly random hybridization temperature (42°C) suggests other potentially novel species may have been missed as probes of varying length mean different optimum hybridization temperatures (Allers *et al*., 2013). The protocol also did not include any step to remove loosely bound probe which increases the probability of false positives. However, the study proved the possibility of not relying on sequence data in designing experiments to identify novel-phage host pairs.

The second reverse metaviromics method involved using phage sequence data to guide the identification and isolation of not only the bacterial host but also infective phages for further experimental characterization. Potentially novel phages, the *Salinivibrio* infecting SMHB1 and Contig13, were identified from the two sites explored (Chapter two). These species were particularly interesting for different reasons. *Salinivibrio* bacterial species thrive in heavy metal rich, high salt environments (Gorriti *et al*., 2014) offering some biotechnological potential. Phage SMHB1 was isolated using the classical approach of culturing bacteria from the same environment and challenging them with the phage fraction. The isolation media and growth conditions were however, selected to target *Salinivibrio*, since *Salinivibrio*-like phage sequences were identified in the metavirome sequence data. As discussed in Chapter three, the integration site, within the tRNA dihydrouridine synthase gene of the host, and the novel nature of the integrase gene suggests that SMHB1 may play some role in the survival of the bacterial

host in such environments (Farrugia *et al*., 2015). Therefore, SMHB1 and other phages that infect these species are worth further research attention.

On the other hand, the attempt to isolate Contig13 phage involved inferring potential bacterial host from sequence data based on similarity to known phages, and challenging the selected bacteria with the environmental phage fraction. Contig13 could not be isolated with the most likely reason being that the phage was not presented with a susceptible host or that the phage particles were no longer infective (Section 2.3.2.5). The similarity to a known *Thalassomonas* phage, BA3, made the screening of *Thalassomonas* bacterial species the logical starting point for host identification but does not necessarily mean the host is a *Thalassomonas* species. Given that through genetic exchange, the swapping of entire modules between phages has been shown to be more prevalent than initially thought (Haggård-Ljungquist *et al*., 1992; Lang *et al*., 2012; Lee *et al*., 2014), therefore a high sequence similarity to a known phage does not necessarily indicate that it infects the same host. The Contig13 phage identified in this study serves to highlight the classic enigma of metaviromic studies and queries the validity of using sequence similarity to predict phage hosts.

At what percentage similarity can we be certain that potentially novel phages would infect hosts of phages to which they are similar? Do we keep screening closely related bacterial species until we find one that is susceptible? To further emphasize this point, the whole genomes of phage SMHB1 and Contig13 were analyzed through HostPhinder, a recently developed bioinformatic tool for phage host prediction (Villarroel *et al*., 2016). The tool analyzes both the query phage sequence and the genomes of described phages with known hosts for co-occurring subsequences (k-mers: DNA sequences of defined length k, in this case 16 nucleotides), and based on factors such as coverage length, host abundance, average accuracy, the bacterial host of the known phage is suggested as the host of the query phage

genome. This is probably the best approach for *in silico* phage host prediction given that the query and reference sequences are broken into overlapping fragments 16 bp long and compared on the basis of co-occurrence rather than location on either genome. However, a number of limitations are obvious. First, a host that has multiple characterized phages, for example, *E. coli*, will be more frequently represented on the reference database, and could be biased for. This cannot be easily solved as it is a consequence of extensive studies of these species. Second, the tool is based on the premise that phages that are genetically similar often infect similar hosts, an assumption which does not always hold true. For example, HostPhinder predicted that the host for phage SMHB1 is either *Vibrio vulnificus* or *Vibrio cholerae*. *V. vulnificus* is the host of phage PV94 to which SMHB1 shares the highest nucleotide similarity while *V. cholera* is the host of phage K139 to which SMHB1 shares the highest protein similarity (Chapter three). These three phages are genetically similar but infect bacterial hosts that differ sufficiently and belong to different genera. SMHB1 might indeed infect either of these bacteria but the widely known specificity of phage infection makes this unlikely. There was no prediction of *Salinivibrio* which this study has shown to be the host of SMHB1. This could be attributed to the fact that SMHB1 and CW02, the only other characterized *Salinivibrio* phage, share no significant nucleotide/protein similarity. The program also predicted that the host for Contig13 is *Thalassomonas loyana* which this study has shown as probably not being the case.

Indeed, studies that have attempted to predict phage hosts solely from sequence data have had mixed results. A recent study analyzed the sequence similarity between characterized phages and bacteria host that have been whole genome sequenced (Edwards *et al*., 2016). The study found that using factors such as genetic homology, CRISPRs or exact matches, the host for only 10-40% of the phages could be correctly identified. The most accurate prediction was when the sequences were screened for exact matches but the authors conceded that a significant proportion of the similarity was due to prophages within the genomes of the bacteria analyzed.

This study reinforced the generally accepted view that there is much insight to be gained from using a bioinformatic approach, but also demonstrated the uncertainty of the correlation between sequence data similarity, host identification and isolation.

All the discussion above can be summarized in one sentence: advances in metaviromics have led to gaps between sequence data generation, phage identification and isolation/characterization. The relatively small size of currently available curated phage databases limits the extent to which the majority of the sequences generated can be annotated. The problem of the large proportion of phage genes being of unknown function is a direct consequence of the high diversity of viral genomes and the rate of discovery of new viral species outstripping their biological characterization (Lima-Mendez *et al*., 2007; Hatfull, 2008). For the portion of metaviromic sequence data that can be annotated, novel phages can be identified and logical extrapolations and conclusions can be drawn from the analysis of those genomes, but ultimately, these have to be validated experimentally. This simply means culturing cannot be completely bypassed if a holistic view of the relationship of a phage with its host is required.

Another reason why host culturing is required is that while some of the novel techniques described in literature and this study have the potential to co-isolate novel phage-host pairs, the methods employed often result in the isolation of non-viable phage particles. For instance, a successful viral tagging experiment will result in a phage attached to the host but which has injected its genetic material and is no longer infective. Similarly, a successful FISH experiment will yield a bacterial host with the phage genome within its cell membrane. Therefore, there is still the need to not only isolate intact phages but also propagate them for further investigation.

Hence, it might be more beneficial to concurrently focus on improving phage isolation and host culturing techniques. This will in turn improve our ability to characterize phage-host pairs, identify the biological functions of currently unknown genes, and improve the available references for the analysis of metaviromic sequence data. This study attempted this approach with the isolation of phage SMHB1, taking into account hits to phage CW02 and growth requirements of previously described *Salinivibrio* species such as salt tolerance and optimum temperature. This led to the discovery of a novel phage and two novel *Salinivibrio* species. Future studies may incorporate bioinformatic analysis of pathways encoded within the bacterial genome, as a measure of identifying compounds the bacteria is likely to metabolize. Such compounds can be incorporated into the growth media to select for the particular host.

Despite the challenges associated with identifying, isolating and characterizing novel phage-host pairs, future studies to address these challenges are absolutely essential for a number of reasons. First, phages encode potentially useful enzymes and studies of such enzymes have led to very interesting discoveries. For example, a phage encoded thermostable DNA polymerase with reverse transcriptase activity was recently identified (Moser *et al*., 2012); phage encoded lytic enzymes have been used to control contaminating *Lactobacillus* in ethanol fermentation (Roach *et al*., 2013); and a phage encoded, colanic acid-degrading enzyme has shown promise in breaking down extracellular matrices of antibiotic resistance pathogenic bacteria (Kim *et al*., 2015). These studies show the importance of investigating the ever evolving interaction between phages and bacteria.

Second, *E. coli* infecting phages such as T4 and Lambda, are the most widely studied (Rabbani *et al*., 2004; Parua *et al*., 2010), of which the latter has been engineered for *in-vitro* packaging of DNA for metagenomic library construction (Sorge, 1988; Lodish *et al*., 2000). The continued study of phage-host interactions is necessary to find another genetic system capable of overcoming the limitations associated with carrying out expression studies in *E. coli.*

Third, the recent discovery of the bacterial CRISPR/*cas* defense mechanism demonstrates the utility of continued study of phage host interactions. It is known that bacteria developed the mechanism to limit the rate of horizontal transfer of genetic elements and to protect against phages and plasmids (H. Deveau *et al*., 2008; Szczepankowska, 2012). The consequence of this constantly evolving evolutionary arm race has been developed into a powerful tool for genome targeting, editing and regulation (Ledford, 2016). The possibilities with this technique are endless, ranging from treating diseases, better understanding of cellular systems to editing the genes of human embryos to remove predisposing factors (Ledford, 2015; Ay, 2016). Phage-host studies therefore still have much to offer to our understanding of biology and in the development of tools for medicine and research.

This study has contributed to the available body of knowledge on the phage community present in an understudied environment, the Namibian Desert. We have shown that there are indeed novel phages in this environment and that high precipitation and fog events may be key to shaping the diversity, an observation that will require further study. We have also described only the second characterized *Salinivibrio* infecting phage, SMHB1. The relatively understudied genus holds some promise in the chemical transformation of heavy metals (Gorriti *et al*., 2014) and it would be interesting to know the role of phages, if any, in that process. The phage-host isolation techniques attempted have not been developed enough to be adopted for the high throughput identification and isolation of novel phages but we have identified some of the limitations associated with each approach and the elements which require improvement.

In general, metaviromics offers a good starting point for rapidly identifying novel phage sequences which can be used with caution to provide valuable insight into phage diversity and abundance in an environment. Factors such as improved virus purification methods, the reducing cost of sequencing, and the relatively small genome size of most phages, are likely to

lead to a rapid increase in the number of metaviromic studies (Culley *et al*., 2006; Delwart, 2007). A large proportion of the generated sequence data is not immediately useful but the situation is expected to improve as more phage genomes are sequenced and, more importantly, biologically characterized.

APPENDICES

APPENDIX A

Accession numbers for the large terminase subunits of phages used for phylogenetic tree construction

Phage	Accession Number
Aeromonas phage Aeh1	NP944105.1
Bacillus phage Bcp1	YP009031337.1
Bacillus phage phiNIT1	YP008318309.1
Bacillus phage SPBc2	NP046608.1
Bacillus phage vB_BanS-Tsamsa	YP008873483.1
Burkholderia phage BEK	AIP84290.1
Burkholderia phage KS14	YP004306879.1
Burkholderia phage phiE12-2	YP001111154.1
Burkholderia phage phiE202	YP001111035.1
Burkholderia prophage phi52237	YP293751.1
Caulobacter vibrioides prophage CC1	WP010920631.1
Clostridium phage c-st	YP398598.1
Deinococcus radiodurans prophage RadMu	WP010889353.1
Escherichia coli phage lambda	NP040581.1
Escherichia coli phage Mu	NP050632.1
Escherichia coli phage P2	NP046758.1
Escherichia coli phage PsP3	NP958058.1

Escherichia coli phage T3	NP523347.1
Escherichia coli phage T4	NP049776.1
Escherichia phage ECBP1	AFR52033.1
Escherichia phage pro483	YP009211911.1
Haemophilus influenza prophage FluMu	WP005693520.1
Haemophilus phage HP1	NP043485.1
Haemophilus phage HP2	NP536821.1
Klebsiella pneumoniae prophage Kpn	SBY00723.1
Mannheimia haemolytica phage M42548	WP015484276.1
Mannheimia phage phiMHaA1	YP655470.1
Mannheimia phage vB_MhM_1152AP	AFL46450.1
Mannheimia phage vB_MhM_587AP1	YP009207755.1
Mycobacterium phage L5	NP039677.1
Pasteurella phage F108	YP654730.1
Pseudomonas phage phiCTX DNA	BAA36228.1,
Pseudomonas phage phiKMV	NP877482.1
Pseudomonas putida phage Θ3	WP010953264.1
Salinivibrio phage CW02	YP007010563.1
Salmonella enterica phage SP6	NP853601.1
Salmonella phage Fels-1	YP001700563.1
Salmonella phage SEN1	YP009217857.1
Salmonella typhimurium phage P22	YP063734.1
Shigella phage SP18	YP003934803.1
Stenotrophomonas phage Smp131	YP009008362.1

Streptomyces phage phiBT1	NP813693.1
Vibrio cholerae phage K139	NP536648.1
Vibrio phage VPUSM 8	YP008766840.1
Vibrio vulnificus Phage PV94	ZP16942787.1
Yersinia phage phiA1122	NP848309.1

APPENDIX B

General buffers and stock solutions used in this study. Solutions marked with an asterisk (*) were autoclaved at 121°C for 20 minutes at 15 psi. Alternatively, solutions/buffers were sterilized using 0.22 µm filters.

50 X TAE buffer (*)

242 g Tris base

18.6 g EDTA

57.1 ml glacial acetic acid

dH_2O to a final volume of 1 L

Ammonium acetate (100mM)

385.4 mg

50 ml of sterile dH_2O

Ampicillin stock

100 mg/ml ampicillin in sterile dH_2O

Stored at -20°C

CTAB/NaCl solution (*)

10% [w/v] CTAB

0.7 M NaCl

DNA loading buffer

20% [v/v] glycerol

0.1 M EDTA

1% [w/v] SDS

0.25% bromophenol blue

Ethidium bromide staining solution

200 mg ethidium bromide powder

20 ml of sterile dH_2O

Sodium chloride solution (5 M) (*)

146.1 g

500 ml of sterile dH_2O

TE buffer, pH 8 (10 mM Tris-HCl; 1 mM) (*)

10 mM Tris-HCl

1 mM EDTA

REFERENCES

Ackermann, H.W. 2007. 5500 Phages examined in the electron microscope. *Archives of Virology*, 152(2): 227–243.

Ackermann, H.W. 2009. Bacteriophages. In M. R. J. Clokie & A. M. Kropinski, eds. *Bacteriophages: Methods and Protocols*. New York: Humana Press: 113–126.

Ackermann, H.W., Tremblay, D. & Moineau, S. 2004. Long-term bacteriophage preservation. *WFCC Newsletter*, 38: 35–40.

Adams, M.H. 1959. *Bacteriophage*. New York: Interscience Publishers Inc.,.

Adams, R. & Burdon, R. 1985. *Molecular Biology of DNA Methylation*. New York, NY: Springer New York.

Adriaenssens, E., van Zyl, L., Cowan, D. & Trindade, M. 2016. Metaviromics of Namib Desert Salt Pans: A Novel Lineage of Haloarchaeal Salterproviruses and a Rich Source of ssDNA Viruses. *Viruses*, 8(1): 14.

Adriaenssens, E.M., Van Zyl, L., De Maayer, P., Rubagotti, E., Rybicki, E., Tuffin, M. & Cowan, D.A. 2015. Metagenomic analysis of the viral community in Namib Desert hypoliths. *Environmental Microbiology*, 17(2): 480–495.

Aggarwal, A.K., Rodgers, D.W., Drottar, M., Ptashne, M. & Harrison, S.C. 1988. Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. *Science (New York, N.Y.)*, 242(4880): 899–907.

Allers, E., Moraru, C., Duhaime, M.B., Beneze, E., Solonenko, N., Barrero-Canosa, J., Amann, R. & Sullivan, M.B. 2013. Single-cell and population level viral infection dynamics revealed by phageFISH, a method to visualize intracellular and free viruses. *Environmental Microbiology*, 15(8): 2306–2318.

Amann, R. & Fuchs, B.M. 2008. Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques. *Nature reviews. Microbiology*, 6(5): 339–48.

Amann, R.I., Binder, B.J., Olson, R.J., Chisholm, S.W., Devereux, R. & Stahl, D.A. 1990. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Applied and Environmental Microbiology*, 56(6): 1919–1925.

Anastasia, A.A. & Rossmann, M.G. 2011. Bacteriophage Assembly. *Viruses*, 3(3): 172.

Andrews-Pfannkoch, C., Fadrosh, D.W., Thorpe, J. & Williamson, S.J. 2010. Hydroxyapatite-mediated separation of double-stranded DNA, single-stranded DNA, and RNA genomes from natural viral assemblages. *Applied and Environmental Microbiology*, 76(15): 5039–5045.

Angly, F., Rodriguez-Brito, B., Bangor, D., McNairnie, P., Breitbart, M., Salamon, P., Felts, B., Nulton, J., Mahaffy, J. & Rohwer, F. 2005. PHACCS, an online tool for estimating the structure and diversity of uncultured viral communities using metagenomic information. *BMC bioinformatics*, 6(1): 41.

Angly, F.E., Felts, B., Breitbart, M., Salamon, P., Edwards, R.A., Carlson, C., Chan, A.M., Haynes, M., Kelley, S., Liu, H., Mahaffy, J.M., Mueller, J.E., Nulton, J., Olson, R., Parsons, R., Rayhawk, S., Suttle, C.A. & Rohwer, F. 2006. The marine viromes of four oceanic regions. *PLoS Biology*, 4(11): 2121–2131.

Angly, F.E., Willner, D., Prieto-Davó, A., Edwards, R.A., Schmieder, R., Vega-Thurber, R., Antonopoulos, D.A., Barott, K., Cottrell, M.T., Desnues, C., Dinsdale, E.A., Furlan, M., Haynes, M., Henn, M.R., Hu, Y., Kirchman, D.L., McDole, T., McPherson, J.D., Meyer, F., Miller, R.M., Mundt, E., Naviaux, R.K., Rodriguez-Mueller, B., Stevens, R., Wegley, L., Zhang, L., Zhu, B. & Rohwer, F. 2009. The GAAS metagenomic tool and its estimations of viral and microbial average genome size in four major biomes. *PLoS computational biology*, 5(12): e1000593.

Anon. 2011. Heavy Rain in Namibia and South Africa : Natural Hazards. http://earthobservatory.nasa.gov/NaturalHazards/view.php?id=50573 24 February 2016.

Arm, J.P., Gurish, M.F., Reynolds, D.S., Scott, H.C., Gartner, C.S., Austen, K.F. & Katz, H.R. 1991. Molecular cloning of gp49, a cell-surface antigen that is preferentially expressed by mouse mast cell progenitors and is a new member of the immunoglobulin superfamily. *Journal of Biological Chemistry*, 266(24): 15966–15973.

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y. & Wishart, D.S. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic acids research*.

Arriola, E., Lambros, M.B.K., Jones, C., Dexter, T., Mackay, A., Tan, D.S.P., Tamber, N., Fenwick, K., Ashworth, A., Dowsett, M. & Reis-Filho, J.S. 2007. Evaluation of Phi29-based whole-genome amplification for microarray-based comparative genomic hybridisation. *Laboratory investigation; a journal of technical methods and pathology*, 87(1): 75–83.

Ashelford, K.E., Day, M.J. & Fry, J.C. 2003. Elevated abundance of bacteriophage infecting bacteria in soil. *Applied and Environmental Microbiology*, 69(1): 285–289.

Ay, H. 2016. Tomorrow's children. *Nature*, 530: 402–405.

Aziz, R.K., Bartels, D., Best, A. a, DeJongh, M., Disz, T., Edwards, R. a, Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R. a, McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A. & Zagnitko, O. 2008. The RAST Server: rapid annotations using subsystems technology. *BMC genomics*, 9: 75.

Bamford, D.H., Palva, E.T. & Lounatmaa, K. 1976. Ultrastructure and life cycle of the lipid-containing bacteriophage phi 6. *The Journal of general virology*, 32(2): 249–59.

Bamford, D.H., Romantschuk, M. & Somerharju, P.J. 1987. Membrane fusion in prokaryotes: bacteriophage phi 6 membrane fuses with the Pseudomonas syringae outer membrane. *The EMBO journal*, 6(5): 1467–73.

Banin, E., Israely, T., Kushmaro, A., Loya, Y., Orr, E. & Rosenberg, E. 2000. Penetration of the coral-bleaching bacterium Vibrio shiloi into Oculina patagonica. *Applied and environmental*

microbiology, 66(7): 3031–6.

Bardarov, S., Bardarov Jr, S., Pavelka Jr, M.S., Sambandamurthy, V., Larsen, M., Tufariello, J., Chan, J., Hatfull, G. & Jacobs Jr, W.R. 2002. Specialized transduction: an efficient method for generating marked and unmarked targeted gene disruptions in Mycobacterium tuberculosis, M. bovis BCG and M. smegmatis. *Microbiology (Reading, England)*, 148(Pt 10): 3007–17.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. & Horvath, P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. , 315(5819): 1709–1712.

Bauer, H., Giebl, H., Hitzenberger, R., Kasper-Giebl, A., Reischl, G., Zibuska, F. & Puxbaum, H. 2003. Airborne bacteria as cloud condensation nuclei. *Journal of Geophysical Research*, 108(D21): 1–5.

Becker, S.C., Dong, S., Baker, J.R., Foster-Frey, J., Pritchard, D.G. & Donovan, D.M. 2009. LysK CHAP endopeptidase domain is required for lysis of live staphylococcal cells. *FEMS Microbiology Letters*, 294(1): 52–60.

Belanger, K.G. & Kreuzer, K.N. 1998. Bacteriophage T4 Initiates Bidirectional DNA Replication through a Two-Step Process. *Molecular Cell*, 2(5): 693–701.

Belcaid, M., Bergeron, A. & Poisson, G. 2011. The evolution of the tape measure protein: units, duplications and losses. *BMC bioinformatics*, 12(Suppl9): S10.

Benhar, I. 2001. Biotechnology applications of phage and cell surface display. *Biotechnol. Adv.*, 19(1): 1–33.

Benzer, S. 1955. Fine Structure of a Genetic Region in Bacteriophage. *Proceedings of the National Academy of Sciences*, 41(6): 344–354.

Bergh, O., Børsheim, K.Y., Bratbak, G. & Heldal, M. 1989. High abundance of viruses found in aquatic environments. *Nature*, 340(6233): 467–8.

Bertani, G. & Weigle, J.J. 1953. Host controlled variation in bacterial viruses. *Journal of bacteriology*, 65(2): 113–121.

Betley, J.N., Frith, M.C., Graber, J.H., Choo, S. & Deshler, J.O. 2002. A ubiquitous and conserved signal for RNA localization in chordates. *Current Biology*, 12(20): 1756–1761.

Bikard, D. & Marraffini, L.A. 2013. Control of gene expression by CRISPR-Cas systems. *F1000Prime reports*, 5(November): 47.

Black, L.W. 1995. DNA packaging and cutting by phage terminases: control in phage T4 by a synaptic mechanism. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 17(12): 1025–30.

Blatny, J.M., Godager, L., Lunde, M. & Nes, I.F. 2004. Complete genome sequence of the Lactococcus lactis temperate phage phiLC3: comparative analysis of phiLC3 and its relatives in lactococci and streptococci. *Virology*, 318(1): 231–244.

Bogovazova, G.G., Voroshilova, N.N. & Bondarenko, V.M. 1991. The efficacy of Klebsiella pneumoniae bacteriophage in the therapy of experimental Klebsiella infection. *Zhurnal mikrobiologii, epidemiologii, i immunobiologii*, (4): 5–8.

Bolotin, A., Quinquis, B., Sorokin, A. & Dusko Ehrlich, S. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. , 151(8): 2551–2561.

Bondy-Denomy, J., Pawluk, A., Maxwell, K.L. & Davidson, A.R. 2013. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*, 493(7432).

Bordet, J. & Ciuca, M. 1921. Remarques sur l'historique de recherches concernant la lyse microbienne transmissible. *Compt. Rend. Soc. Biol*, 84: 745–747.

Borysowski, J., Weber-Dabrowska, B. & Górski, A. 2006. Bacteriophage endolysins as a novel class of antibacterial agents. *Experimental Biology and Medicine*, 231(1535-3702 (Print)): 366–377.

Botstein, D. 1980. A theory of modular evolution for bacteriophages. *Annals of the New York Academy*

of Sciences, 354: 484–90.

Boujelben, I., Yarza, P., Almansa, C., Villamor, J., Maalej, S., Antón, J. & Santos, F. 2012. Virioplankton community structure in Tunisian solar salterns. *Applied and environmental microbiology*, 78(20): 7429–37.

Boyd, E.F., Almagro-Moreno, S. & Parent, M.A. 2009. Genomic islands are dynamic, ancient integrative elements in bacterial evolution. *Trends in microbiology*, 17(2): 47–53.

Bradford, M.M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, 72(1-2): 248–254.

Breitbart, M., Miyake, J.H. & Rohwer, F. 2004. Global distribution of nearly identical phage-encoded DNA sequences. *FEMS Microbiology Letters*, 236(2): 249–256.

Brister, J.R. 2008. Origin Activation Requires both Replicative and Accessory Helicases during T4 Infection. *Journal of Molecular Biology*, 377(5): 1304–1313.

Brister, J.R. & Nossal, N.G. 2007. Multiple Origins of Replication Contribute to a Discontinuous Pattern of DNA Synthesis Across the T4 Genome During Infection. *Journal of Molecular Biology*, 368(2): 336–348.

Brown, J.E., Klement, J.F. & Mcallister, W.T. 1986. Sequences of three promoters for the bacteriophage SP6 RNA polymerase. *Nucleic Acids Research*, 14(8): 3521–3526.

Brum, J., Steward, G., Jiang, S. & Jellison, R. 2005. Spatial and temporal variability of prokaryotes, viruses, and viral infections of prokaryotes in an alkaline, hypersaline lake. *Aquatic Microbial Ecology*, 41: 247–260.

Brum, J.R., Schenck, R.O. & Sullivan, M.B. 2013. Global morphological analysis of marine viruses shows minimal regional variation and dominance of non-tailed viruses. *The International Society for Microbial Ecology journal*, 7(9): 1738–51.

Brussaard, C.P.D. 2004. Optimization of Procedures for Counting Viruses by Flow Cytometry. *Applied*

and Environmental Microbiology, 70(3): 1506–1513.

Brussaard, C.P.D., Marie, D. & Bratbak, G. 2000. Flow cytometric detection of viruses. *Journal of Virological Methods*, 85(1-2): 175–182.

Brüssow, H., Canchaya, C. & Hardt, W.-D. 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiology and molecular biology reviews : MMBR*, 68(3): 560–602.

Brüssow, H. & Desiere, F. 2001. Comparative phage genomics and the evolution of siphoviridae: Insights from dairy phages. *Molecular Microbiology*, 39(2): 213–222.

Büdel, B., Darienko, T., Deutschewitz, K., Dojani, S., Friedl, T., Mohr, K.I., Salisch, M., Reisser, W. & Weber, B. 2009. Southern african biological soil crusts are ubiquitous and highly diverse in drylands, being restricted by rainfall frequency. *Microbial Ecology*, 57(2): 229–247.

Burke, A. 2001. Determining landscape function and ecosystem dynamics: contribution to ecological restoration in the southern Namib desert. *Ambio*, 30(1): 29–36.

Butcher, S.J., Bamford, D.H. & Fuller, S.D. 1995. DNA packaging orders the membrane of bacteriophage PRD1. *The EMBO journal*, 14(24): 6078–86.

Cairns, J., Stent, G.S. & Watson, J.D. 2007. *Phage and the Origins of Molecular Biology*. Plainview, N.Y.: Cold Spring Harbor Laboratory Press.

Campanizzi, J. 2014. Ptashne reflects on the chemistry of gene regulation. *American Society for Biochemistry and Molecular Biology*. http://www.asbmb.org/asbmbtoday/201404/JournalNews/JBC/Ptashne/ 31 March 2016.

Campoy, S., Aranda, J., Àlvarez, G., Barbé, J. & Llagostera, M. 2006. Isolation and sequencing of a temperate transducing phage for Pasteurella multocida. *Applied and Environmental Microbiology*, 72(5): 3154–3160.

Canchaya, C., Fournous, G. & Brüssow, H. 2004. The impact of prophages on bacterial chromosomes.

Molecular Microbiology, 53(1): 9–18.

Canchaya, C., Proux, C., Fournous, G., Bruttin, A. & Brüssow, H. 2003. Prophage genomics. *Microbiology and molecular biology reviews : MMBR*, 67(2): 238–276.

Cann, A.J. 1994. *Principles of molecular virology*. 2nd ed. Academic Press.

Carles-Kinch, K. & Kreuzer, K.N. 1997. RNA-DNA hybrid formation at a bacteriophage T4 replication origin. *Journal of Molecular Biology*, 266(5): 915–926.

Carlton, R.M. 1999. Phage therapy: past history and future prospects. *Archivum immunologiae et therapiae experimentalis*, 47(5): 267–274.

Casjens, S., Hatfull, G. & Hendrix, R. 1992. Evolution of dsDNA tailed-bacteriophage genomes. *Seminars in Virology*, 3: 383–397.

Casjens, S.R. 2005. Comparative genomics and evolution of the tailed-bacteriophages. *Current Opinion in Microbiology*, 8(4): 451–458.

Cassuto, E., Lash, T., Sriprakash, K.S. & Radding, C.M. 1971. Role of exonuclease and beta protein of phage lambda in genetic recombination. V. Recombination of lambda DNA in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, 68(7): 1639–1643.

Cerritelli, M.E., Wall, J.S., Simon, M.N., Conway, J.F. & Steven, A.C. 1996. Stoichiometry and Domainal Organization of the Long Tail-fiber of Bacteriophage T4: A Hinged Viral Adhesin. *Journal of Molecular Biology*, 260(5): 767–780.

Chatterjee, S. & Rothenberg, E. 2012. Interaction of bacteriophage l with its E. coli receptor, LamB. *Viruses*, 4(11): 3162–78.

Chaturongakul, S. & Ounjai, P. 2014. Phage-host interplay: examples from tailed phages and Gram-negative bacterial pathogens. *Frontiers in microbiology*, 5: 442.

Chen, F. & Lu, J. 2002. Genomic sequence and evolution of marine cyanophage P60: A new insight on lytic and lysogenic phages. *Applied and Environmental Microbiology*, 68(5): 2589–2594.

Cheng, X. & Patterson, T.A. 1992. Construction and use of lambda PL promoter vectors for direct cloning and high level expression of PCR amplified DNA coding sequences. *Nucleic Acids Research*, 20(17): 4591–4598.

Chibani-chennoufi, S., Bruttin, A., Brüssow, H., Dillmann, M. & Bru, H. 2004. Phage-Host Interaction : an Ecological Perspective. *Journal of bacteriology*, 186(12): 3677–3686.

Chopin, M.C., Chopin, A. & Bidnenko, E. 2005. Phage abortive infection in lactococci: Variations on a theme. *Current Opinion in Microbiology*, 8(4): 473–479.

Chung, W.J., Lee, D.Y. & Yoo, S.Y. 2014. Chemical modulation of m13 bacteriophage and its functional opportunities for nanomedicine. *International Journal of Nanomedicine*, 9(1): 5825–5836.

Clark, A.J., Inwood, W., Cloutier, T. & Dhillon, T.S. 2001. Nucleotide sequence of coliphage HK620 and the evolution of lambdoid phages. *Journal of molecular biology*, 311(4): 657–679.

Clark, J.R. & March, J.B. 2006. Bacteriophages and biotechnology: vaccines, gene therapy and antibacterials. *Trends in Biotechnology*, 24(5): 212–218.

Claverie, J.M., Ogata, H., Audic, S., Abergel, C., Suhre, K. & Fournier, P.E. 2006. Mimivirus and the emerging concept of 'giant' virus. *Virus Research*, 117(1): 133–144.

Click, E.M. & Webster, R.E. 1998. The TolQRA Proteins Are Required for Membrane Insertion of the Major Capsid Protein of the Filamentous Phage f1 during Infection. *Journal of Bacteriology*, 180(7): 1723–1728.

Clokie, M.R., Millard, A.D., Letarov, A. V & Heaphy, S. 2011. Phages in nature. *Bacteriophage*, 1(1): 31–45.

Clutter, M.R., Heffner, G.C., Krutzik, P.O., Sachen, K.L. & Nolan, G.P. 2010. Tyramide signal amplification for analysis of kinase activity by intracellular flow cytometry. *Cytometry Part A*, 77(11): 1020–1031.

Cochran, P.K. & Paul, J.H. 1998. Seasonal abundance of lysogenic bacteria in a subtropical estuary. *Applied and Environmental Microbiology*, 64(6): 2308–2312.

Collins, J. & Hohn, B. 1978. Cosmids: a type of plasmid gene-cloning vector that is packageable in vitro in bacteriophage lambda heads. *Proceedings of the National Academy of Sciences of the United States of America*, 75(9): 4242–4246.

Comeau, A.M., Bertrand, C., Letarov, A., Tétart, F. & Krisch, H.M. 2007. Modular architecture of the T4 phage superfamily: A conserved core genome and a plastic periphery. *Virology*, 362(2): 384–396.

Conley, M.P. & Wood, W.B. 1975. Bacteriophage T4 whiskers: a rudimentary environment-sensing device. *Proc Natl Acad Sci U S A*, 72(9): 3701–3705.

Coombs, D.H. & Eiserling, F.A. 1977. Studies on the structure, protein composition and assembly of the neck of bacteriophage T4. *Journal of Molecular Biology*, 116(3): 375–405.

Court, D.L., Oppenheim, A.B. & Adhya, S.L. 2007. A New Look at Bacteriophage λ Genetic Networks. *Journal of Bacteriology*, 189(2): 298–304.

Cresawn, S.G., Bogel, M., Day, N., Jacobs-Sera, D., Hendrix, R.W. & Hatfull, G.F. 2011. Phamerator: a bioinformatic tool for comparative bacteriophage genomics. *BMC Bioinformatics*, 12(1): 395.

Culley, A., Lang, A. & Suttle, C. 2006. Metagenomic analysis of coastal RNA virus communities. *Science*, 312(5781): 1795–1798.

Culley, A.I., Lang, A.S. & Suttle, C.A. 2006. Metagenomic analysis of coastal RNA virus communities. *Science (New York, N.Y.)*, 312(5781): 1795–8.

Czyz, A., Los, M., Wrobel, B. & Wegrzyn, G. 2001. Inhibition of spontaneous induction of lambdoid prophages in Escherichia coli cultures: simple procedures with possible biotechnological applications. *BMC biotechnology*, 1: 1.

d'Herelle, F. 1917. Sur un microbe invisible antagonistedes bacilles dysenteriques. *Les Comptes rendus*

del'Académie des sciences, 165: 373–5.

Daims, H., Brühl, A., Amann, R., Schleifer, K.H. & Wagner, M. 1999. The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set. *Systematic and applied microbiology*, 22(3): 434–44.

Dale, J. 2004. *Molecular Genetics of Bacteria*. 3rd ed. Washington D.C.: ASM Press.

Dalluge, J.J., Hashizume, T., Sopchik, A.E., McCloskey, J.A. & Davis, D.R. 1996. Conformational flexibility in RNA: The role of dihydrouridine. *Nucleic Acids Research*, 24(6): 1073–1079.

Danovaro, R. & Serresi, M. 2000. Viral density and virus-to-bacterium ratio in deep-sea sediments of the Eastern Mediterranean. *Applied and Environmental Microbiology*, 66(5): 1857–1861.

Darzentas, N. 2010. Circoletto: Visualizing sequence similarity with Circos. *Bioinformatics*, 26(20): 2620–2621.

Day, J.A. 1993. The major ion chemistry of some southern African saline systems. *Hydrobiologia*, 267(1-3): 37–59.

Day, J.A. & Seely, M.K. 1988. Physical and chemical conditions in an hypersaline spring in the Namib Desert. *Hydrobiologia*, 160(2): 141–153.

DeLeon-Rodriguez, N., Lathem, T.L., Rodriguez-R, L.M., Barazesh, J.M., Anderson, B.E., Beyersdorf, A.J., Ziemba, L.D., Bergin, M., Nenes, A. & Konstantinidis, K.T. 2013. Microbiome of the upper troposphere: species composition and prevalence, effects of tropical storms, and atmospheric implications. *Proceedings of the National Academy of Sciences of the United States of America*, 110(7): 2575–80.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J. & Charpentier, E. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. , 471(7340): 602–607.

Delwart, E.L. 2007. Viral metagenomics. *Reviews in Medical Virology*, 17(2): 115–131.

Deng, L., Gregory, A., Yilmaz, S., Poulos, B.T., Hugenholtz, P. & Sullivan, M.B. 2012. Contrasting life strategies of viruses that infect photo- and heterotrophic bacteria, as revealed by viral tagging. *mBio*, 3(6): 1–8.

Dernburg, A.F., Broman, K.W., Fung, J.C., Marshall, W.F., Philips, J., Agard, D.A. & Sedat, J.W. 1996. Perturbation of nuclear architecture by long-distance chromosome interactions. *Cell*, 85(5): 745–759.

Deschavanne, P., DuBow, M.S. & Regeard, C. 2010. The use of genomic signature distance between bacteriophages and their hosts displays evolutionary relationships and phage growth cycle determination. *Virology journal*, 7: 163.

Desnues, C., Rodriguez-Brito, B., Rayhawk, S., Kelley, S., Tran, T., Haynes, M., Liu, H., Furlan, M., Wegley, L., Chau, B., Ruan, Y., Hall, D., Angly, F.E., Edwards, R.A., Li, L., Thurber, R.V., Reid, R.P., Siefert, J., Souza, V., Valentine, D.L., Swan, B.K., Breitbart, M. & Rohwer, F. 2011. Biodiversity and Biogeography of Phages in Modern Stromatolites and Thrombolites. *Handbook of Molecular Microbial Ecology II: Metagenomics in Different Habitats*, 452(7185): 37–44.

Deveau, H., Barrangou, R., Garneau, J.E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P. & Moineau, S. 2008. Phage Response to CRISPR-Encoded Resistance in Streptococcus thermophilus. *Journal of Bacteriology*, 190(4): 1390–1400.

Deveau, H., Barrangou, R., Garneau, J.E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P. & Moineau, S. 2008. Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *Journal of Bacteriology*, 190(4): 1390–1400.

Dickson, R.C., Barnes, S.L. & Eiserling, F.A. 1970. Structural proteins of bacteriophage T4. *Journal of Molecular Biology*, 53(3): 461–474.

Diemer, G.S. & Stedman, K.M. 2012. A novel virus genome discovered in an extreme environment suggests recombination between unrelated groups of RNA and DNA viruses. *Biology direct*, 7: 13.

Dobbins, A.T., M. Jr, G., Basham, D.A., Ford, M.E., Houtz, J.M., Pedulla, M.L., Lawrence, J.G., Hatfull, G.F. & Hendrix, R.W. 2004. Complete genomic sequence of the virulent *Salmonella* bacteriophage SP6. *Journal of Bacteriology*, 186(7): 1933–1944.

Dokland, T. & Murialdo, H. 1993. Structural transitions during maturation of bacteriophage lambda capsids. *Journal of Molecular Biology*, 233(4): 682–694.

Dotto, G.P., Horiuchi, K. & Zinder, N.D. 1984. The functional origin of bacteriophage f1 DNA replication. Its signals and domains. *Journal of Molecular Biology*, 172(4): 507–521.

Dreiseikelmann, B. 1994. Translocation of DNA across bacterial membranes. *Microbiological reviews*, 58(3): 293–316.

Drulis-Kawa, Z., Majkowska-Skrobek, G., Maciejewska, B., Delattre, A.-S. & Lavigne, R. 2012. Learning from bacteriophages - advantages and limitations of phage and phage-encoded protein applications. *Current protein & peptide science*, 13(8): 699–722.

Dueker, M.E., O'Mullan, G.D., Weathers, K.C., Juhl, A.R. & Uriarte, M. 2012. Coupling of fog and marine microbial content in the near-shore coastal environment. *Biogeosciences*, 9(2): 803–813.

van Duin, J. 1988. The single-stranded RNA bacteriophages. In R. Calendar, ed. *The Bacteriophages*. New York: Plenum: 117–67.

Dy, R.L., Przybilski, R., Semeijn, K., Salmond, G.P.C. & Fineran, P.C. 2014. A widespread bacteriophage abortive infection system functions through a Type IV toxin-antitoxin mechanism. *Nucleic acids research*, 42(7): 4590–605.

Echols, H. & Court, D. 1971. The role of helper phage in gal transduction. In A. Hershey, ed. *The Bacteriophage Lambda*. New York: Cold Spring Harbor Press: 701–710.

Eckardt, F.D. & Drake, N. 2011. *Sabkha Ecosystems*. M. Öztürk, B. Böer, H. J. Barth, M. Clüsener-Godt, M. A. Khan, & S. W. Breckle, eds. Dordrecht: Springer Netherlands.

Eckardt, F.D., Drake, N., Goudie, A.S., White, K. & Viles, H. 2001. The role of playas in pedogenic

gypsum crust formation in the Central Namib Desert: a theoretical model. *Earth Surface Processes and Landforms*, 26(11): 1177–1193.

Eckardt, F.D., Soderberg, K., Coop, L.J., Muller, A.A., Vickery, K.J., Grandin, R.D., Jack, C., Kapalanga, T.S. & Henschel, J. 2013. The nature of moisture at Gobabeb, in the central Namib Desert. *Journal of Arid Environments*, 93: 7–19.

Edgar, R., McKinstry, M., Hwang, J., Oppenheim, A.B., Fekete, R. a, Giulian, G., Merril, C., Nagashima, K. & Adhya, S. 2006. High-sensitivity bacterial detection using biotin-tagged phage and quantum-dot nanocomplexes. *Proceedings of the National Academy of Sciences of the United States of America*, 103(17): 4841–5.

Edgar, R.C. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics*, 5: 113.

Edwards, R.A., McNair, K., Faust, K., Raes, J. & Dutilh, B.E. 2016. Computational approaches to predict bacteriophage-host relationships. *FEMS Microbiology Reviews*, 40(2): 258–272.

Edwards, R.A. & Rohwer, F. 2005. Opinion: Viral metagenomics. *Nature Reviews Microbiology*, 3(6): 504–510.

Efrony, R., Atad, I. & Rosenberg, E. 2009. Phage therapy of coral white plague disease: Properties of phage BA3. *Current Microbiology*, 58(2): 139–145.

Egley, L.E.W. & Breitbart, M. 2003. Use of Fluorescently Labeled Phage in the Detection and Identi cation of Bacterial Species. *Applied Spectroscopy*, 57(9): 7.

Emerson, J.B., Thomas, B.C., Andrade, K., Allen, E.E., Heidelberg, K.B. & Banfielda, J.F. 2012. Dynamic viral populations in hypersaline systems as revealed by metagenomic assembly. *Applied and Environmental Microbiology*, 78(17): 6309–6320.

Endemann, H. & Model, P. 1995. Location of filamentous phage minor coat proteins in phage and in infected cells. *Journal of molecular biology*, 250(4): 496–506.

Esposito, D., Fitzmaurice, W.P., Benjamin, R.C., Goodman, S.D., Waldman, A.S. & Scocca, J.J. 1996. The complete nucleotide sequence of bacteriophage HP1 DNA. *Nucleic Acids Research*, 24(12): 2360–2368.

Farrelly, V., Rainey, F.A. & Stackebrandt, E. 1995. Effect of genome size and rrn gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. *Applied and Environmental Microbiology*, 61(7): 2798–2801.

Farrugia, D.N., Elbourne, L.D.H., Mabbutt, B.C. & Paulsen, I.T. 2015. A novel family of integrases associated with prophages and genomic islands integrated within the tRNA-dihydrouridine synthase A (dusA) gene. *Nucleic acids research*, 43(9): 4547–57.

Feiner, R., Argov, T., Rabinovich, L., Sigal, N., Borovok, I. & Herskovits, A.A. 2015. A new perspective on lysogeny: prophages as active regulatory switches of bacteria. *Nature Reviews Microbiology*, 13(10): 641–650.

Feiss, M. & Catalano, C.E. 2005. Bacteriophage lambda terminase and the mechanism of viral DNA packaging. *Viral Genome Packaging Machines: Genetics, Structure, and Mechanism*: 5–39.

Femino, A.M. 1998. Visualization of Single RNA Transcripts in Situ. *Science*, 280(5363): 585–590.

Feucht, A., Schmid, A., Benz, R., Schwarz, H. & Heller, K.J. 1990. Pore formation associated with the tail-tip protein pb2 of bacteriophage T5. *The Journal of biological chemistry*, 265(30): 18561–7.

Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., Min Jou, W., Molemans, F., Raeymaekers, a, Van den Berghe, a, Volckaert, G. & Ysebaert, M. 1976. Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature*, 260(5551): 500–507.

Fink, B., Loepfe, E. & Wyler, R. 1979. Demonstration of viral antigen in cryostat sections by a new immunoperoxidase procedure eliminating endogenous peroxidase activity. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society*, 27(2): 686–8.

Fischetti, V.A. 2010. Bacteriophage endolysins: A novel anti-infective to control Gram-positive

pathogens. *International Journal of Medical Microbiology*, 300(6): 357–362.

Fokine, A., Zhang, Z., Kanamaru, S., Bowman, V.D., Aksyuk, A.A., Arisaka, F., Rao, V.B. & Rossmann, M.G. 2013. The Molecular Architecture of the Bacteriophage T4 Neck. *Journal of Molecular Biology*, 425(10): 1731–1744.

Fonnesbech Vogel, B., Venkateswaran, K., Satomi, M. & Gram, L. 2005. Identification of Shewanella baltica as the most important H2S-producing species during iced storage of Danish marine fish. *Applied and environmental microbiology*, 71(11): 6689–97.

Ford, M.E., Stenstrom, C., Hendrix, R.W. & Hatfull, G.F. 1998. Mycobacteriophage TM4: genome structure and gene expression. *Tubercle and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease*, 79(2): 63–73.

Frederickson, C.M., Short, S.M. & Suttle, C.A. 2003. The Physical Environment Affects Cyanophage Communities in British Columbia Inlets. *Microbial Ecology*, 46(3): 348–357.

Frossard, A., Ramond, J.-B., Seely, M. & Cowan, D.A. 2015. Water regime history drives responses of soil Namib Desert microbial communities to wetting events. *Scientific reports*, 5: 12263.

Fuchs, B.M., Pernthaler, J., Beveridge, T.J., Breznak, a, Marzluf, G., Schmidt, T.M., Amann, R., Ludwig, W., Wagner, M., Sattler, B., Wille, a, Alfreider, a, Psenner, R. & Trebesius, K. 2010. Single cell identification by in situ hybridization. *Methods for General and Molecular Microbiology*, 3(October): 886–896.

Fuhrman, J.A. 1999. Marine viruses and their biogeochemical and ecological effects. *Nature*, 399(6736): 541–8.

Funatsu, T., Taniyama, T., Tajima, T., Tadakuma, H. & Namiki, H. 2002. Rapid and sensitive detection method of a bacterium by using a GFP reporter phage. *Microbiology and immunology*, 46(6): 365–369.

Gabashvili, I.S., Khan, S. a, Hayes, S.J. & Serwer, P. 1997. Polymorphism of bacteriophage T7. *Journal of molecular biology*, 273(3): 658–67.

Garcia-Heredia, I., Martin-Cuadrado, A.-B., Mojica, F.J.M., Santos, F., Mira, A., Antón, J. & Rodriguez-Valera, F. 2012. Reconstructing viral genomes from the environment using fosmid clones: the case of haloviruses. *PloS one*, 7(3): e33802.

Gardner, M., Bann, D., Wiley, L., Cooper, R., Hardy, R., Nitsch, D., Martin-Ruiz, C., Shiels, P., Sayer, A.A., Barbieri, M., Bekaert, S., Bischoff, C., Brooks-Wilson, A., Chen, W., Cooper, C., Christensen, K., De Meyer, T., Deary, I., Der, G., Roux, A.D., Fitzpatrick, A., Hajat, A., Halaschek-Wiener, J., Harris, S., Hunt, S.C., Jagger, C., Jeon, H.S., Kaplan, R., Kimura, M., Lansdorp, P., Li, C., Maeda, T., Mangino, M., Nawrot, T.S., Nilsson, P., Nordfjall, K., Paolisso, G., Ren, F., Riabowol, K., Robertson, T., Roos, G., Staessen, J.A., Spector, T., Tang, N., Unryn, B., van der Harst, P., Woo, J., Xing, C., Yadegarfar, M.E., Park, J.Y., Young, N., Kuh, D., von Zglinicki, T. & Ben-Shlomo, Y. 2014. Gender and telomere length: Systematic review and meta-analysis. *Experimental Gerontology*, 51(1): 15–27.

Garneau, J.E., Dupuis, M.È., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H. & Moineau, S. 2010. The CRISPR/cas bacterial immune system cleaves bacteriophage and plasmid DNA. , 468(7320): 67–71.

Gautheret, D. & Lambert, A. 2001. Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. *Journal of molecular biology*, 313(5): 1003–1011.

Ghannad, M.S. & Mohammadi, A. 2012. Bacteriophage: Time to re-evaluate the potential of phage therapy as a promising agent to control multidrug-resistant bacteria. *Iranian Journal of Basic Medical Sciences*, 15(2): 693–701.

Gitlin, G., Bayer, E.A. & Wilchek, M. 1987. Studies on the biotin-binding site of avidin. Lysine residues involved in the active site. *The Biochemical journal*, 242(3): 923–6.

Goel, U., Kauri, T., Kushner, D.J. & Ackermann, H.-W. 1996. from a Spanish saltern and its lytic bacteriophage. *Canadian Journal of Microbiology*, 42(10): 1015–1023.

Goldberg, G.W., Jiang, W., Bikard, D. & Marraffini, L.A. 2014. Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. , 514(7524): 633–637.

Goldstein, M., Syvanen, M. & Lin, R. 1984. *Bacteria, Plasmids, and Phages: An introduction to Molecular Biology*. Harvard University Press. Cambridge, Massachusetts, and London England.

Goodridge, L., Chen, J. & Griffiths, M. 1999a. Development and characterization of a fluorescent-bacteriophage assay for detection of Escherichia coli O157:H7. *Applied and environmental microbiology*, 65(4): 1397–404.

Goodridge, L., Chen, J. & Griffiths, M. 1999b. The use of a fluorescent bacteriophage assay for detection of Escherichia coli O157:H7 in inoculated ground beef and raw milk. *International journal of food microbiology*, 47(1-2): 43–50.

Gorriti, M.F., Dias, G.M., Chimetto, L.A., Trindade-Silva, A.E., Silva, B.S., Mesquita, M.M.A., Gregoracci, G.B., Farias, M.E., Thompson, C.C. & Thompson, F.L. 2014. Genomic and phenotypic attributes of novel salinivibrios from stromatolites, sediment and water from a high altitude lake. *BMC genomics*, 15(1): 473.

Gottesman, M.E. & Yarmolinsky, M.B. 1968. Integration-negative mutants of bacteriophage lambda. *Journal of Molecular Biology*, 31(3): 487–505.

Grissa, I., Vergnaud, G. & Pourcel, C. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic acids research*, 35(Web Server issue): W52–7.

de Haas, R.R., Verwoerd, N.P., van der Corput, M.P., van Gijlswijk, R.P., Siitari, H. & Tanke, H.J. 1996. The use of peroxidase-mediated deposition of biotin-tyramide in combination with time-resolved fluorescence imaging of europium chelate label in immunohistochemistry and in situ hybridization. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society*, 44(10): 1091–9.

Haggard-Ljungquist, E., Halling, C. & Calendar, R. 1992. DNA sequences of the tail fiber genes of bacteriophage P2: Evidence for horizontal transfer of tail fiber genes among unrelated

bacteriophages. *Journal of Bacteriology*, 174(5): 1462–1477.

Haggård-Ljungquist, E., Halling, C. & Calendar, R. 1992. DNA sequences of the tail fiber genes of bacteriophage P2: evidence for horizontal transfer of tail fiber genes among unrelated bacteriophages. *Journal of bacteriology*, 174(5): 1462–77.

Haggard-Ljungquist, E., Jacobsen, E., Rishovd, S., Six, E.W., Nilssen, O., Sunshine, M.G., Lindqvist, B.H., Kim, K.J., Barreiro, V., Koonin, E. V & Et Al. 1995. Bacteriophage P2: genes involved in baseplate assembly. *Virology*, 213(1): 109–121.

Hambly, E. & Suttle, C.A. 2005. The viriosphere, diversity, and genetic exchange within phage communities. *Current Opinion in Microbiology*, 8(4): 444–450.

Hankin, E.H. 1896. L'action bactericide des eaux de la Jumna et du Gange sur le vibrion du cholera. *Annales de l'Institut Pasteur*, 10: 511–523.

Hatfull, G.F. 2008. Bacteriophage genomics. *Current Opinion in Microbiology*, 11(5): 447–453.

Hatfull, G.F., Pedulla, M.L., Jacobs-Sera, D., Cichon, P.M., Foley, A., Ford, M.E., Gonda, R.M., Houtz, J.M., Hryckowian, A.J., Kelchner, V.A., Namburi, S., Pajcini, K. V., Popovich, M.G., Schleicher, D.T., Simanek, B.Z., Smith, A.L., Zdanowicz, G.M., Kumar, V., Peebles, C.L., Jacobs, W.R., Lawrence, J.G. & Hendrix, R.W. 2006. Exploring the mycobacteriophage metaproteome: Phage genomics as an educational platform. *PLoS Genetics*, 2(6): 0835–0847.

Hatfull, G.F. & Sarkis, G.J. 1993. DNA sequence, structure and gene expression of mycobacteriophage L5: A phage system for mycobacterial genetics. *Molecular Microbiology*, 7(3): 395–405.

Hendrix, R.W. 2002. Bacteriophages: Evolution of the Majority. *Theoretical Population Biology*, 61(4): 471–480.

Hendrix, R.W., Lawrence, J.G., Hatfull, G.F. & Casjens, S. 2000. The origins and ongoing evolution of viruses. *Trends in Microbiology*, 8(11): 504–508.

Hendrix, R.W., Smith, M.C., Burns, R.N., Ford, M.E. & Hatfull, G.F. 1999. Evolutionary relationships

among diverse bacteriophages and prophages: all the world's a phage. *Proceedings of the National Academy of Sciences of the United States of America*, 96(5): 2192–7.

Hennes, K.P., Suttle, C.A. & Chan, A.M. 1995. Fluorescently labeled virus probes show that natural virus populations can control the structure of marine microbial communities. *Applied and Environmental Microbiology*, 61(10): 3623–3627.

Henschel, J.R. & Lancaster, N. 2013. Gobabeb - 50 years of Namib Desert research. *Journal of Arid Environments*, 93: 1–6.

Hershey, A.D. 1952. Independent Functions of Viral Protein and Nucleic Acid in Growth of Bacteriophage. *Journal of General Physiology*, 36(1): 39–56.

Herzenberg, L.A., Parks, D., Sahaf, B., Perez, O., Roederer, M. & Herzenberg, L.A. 2002. The history and future of the Fluorescence Activated Cell Sorter and flow cytometry: A view from Stanford. *Clinical Chemistry*, 48(10): 1819–1827.

Hesse, U., van Heusden, P., Kirby, B., Olonade, I., van Zyl, L.J. & Trindade, M.I. 2016. Metavirome assembly and annotation: a surprise in the Namib Desert.

Hewson, I., O'Neil, J.M., Fuhrman, J. a. & Dennison, W.C. 2001. Virus-like particle distribution and abundance in sediments and overlying waters along eutrophication gradients in two subtropical estuaries. *Limnology and Oceanography*, 46(7): 1734–1746.

Hill, R.W., White, B.A., Cottrell, M.T. & Dacey, J.W.H. 1998. Virus-mediated total release of dimethylsulfoniopropionate from marine phytoplankton: A potential climate process. *Aquatic Microbial Ecology*, 14(1): 1–6.

Hinton, D.M. & Nossal, N.G. 1987. Bacteriophage T4 DNA primase-helicase. Characterization of oligomer synthesis by T4 61 protein alone and in conjunction with T4 41 protein. *The Journal of biological chemistry*, 262(22): 10873–8.

Hinton, D.M., Stitt, B., Hinton, D., Hinton, D., Pande, S., Wais, N., Johnson, X., Vuthoori, M., Makela, A., Hook-Barnard, I., Brody, E., Rabussay, D., Hall, D., Miller, E., Kutter, E., Mosig, G., Arisaka,

F., Kunisawa, T., Ruger, W., Wilkens, K., Ruger, W., Weisberg, R., Hinton, D., Adhya, S., Hinton, D., Schmidt, R., Kreuzer, K., Stevens, A., Stevens, A., Ouhammouch, M., Orsini, G., Brody, E., Ouhammouch, M., Adelman, K., Harvey, S., Orsini, G., Brody, E., Hinton, D., March-Amegadzie, R., Gerber, J., Sharma, M., Browning, D., Busby, S., Murakami, K., Darst, S., Young, B., Gruber, T., Gross, C., Paget, M., Helmann, J., Gruber, T., Gross, C., Campbell, E., Westblade, L., Darst, S., Hook-Barnard, I., Hinton, D., Helmann, J., Vassylyev, D., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M., Borukhov, S., Yokoyama, S., Murakami, K., Masuda, S., Campbell, E., Muzzin, O., Darst, S., Murakami, K., Masuda, S., Darst, S., Campbell, E., Muzzin, O., Chlenov, M., Sun, J., Olson, C., Weinman, O., Trester-Zedlitz, M., Darst, S., Lambert, L., Wei, Y., Schirf, V., Demeler, B., Werner, M., Jain, D., Nickels, B., Sun, L., Hochschild, A., Darst, S., Vassylyev, D., Vassylyeva, M., Perederina, A., Tahirov, T., Artsimovitch, I., Malhotra, A., Severinova, E., Darst, S., Patikoglou, G., Westblade, L., Campbell, E., Lamour, V., Lane, W., Darst, S., Sasse-Dwight, S., Gralla, J., Kontur, W., Saecker, R., Davis, C., Capp, M., Record, M., Saecker, R., Tsodikov, O., McQuade, K., Schlax, P., Capp, M., Record, M., Erie, D., Landick, R., Ciampi, M., Gilmour, D., Fan, R., Adhya, S., Gottesman, M., Wilkens, K., Ruger, W., Koch, T., Raudonikiene, A., Wilkens, K., Ruger, W., Horvitz, H., Horvitz, H., Depping, R., Lohaus, C., Meyer, H., Ruger, W., Ross, W., Gosink, K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K., Gourse, R., Gaal, T., Ross, W., Blatter, E., Tang, H., Jia, X., Krishnan, V., Assa-Munt, N., Ebright, R., Gourse, R., Murakami, K., Fujita, N., Ishihama, A., Zou, C., Fujita, N., Igarashi, K., Ishihama, A., Tiemann, B., Depping, R., Gineikiene, E., Kaliniene, L., Nivinskas, R., Ruger, W., Wilkens, K., Tiemann, B., Bazan, F., Ruger, W., Sommer, N., Salniene, V., Gineikiene, E., Nivinskas, R., Ruger, W., Kashlev, M., Nudler, E., Goldfarb, A., White, T., Kutter, E., Severinov, K., Kashlev, M., Severinova, E., Bass, I., McWilliams, K., Kutter, E., Nikiforov, V., Snyder, L., Goldfarb, A., Tiemann, B., Depping, R., Ruger, W., Pene, C., Uzan, M., Sanson, B., Uzan, M., Hirano, N., Ohshima, H., Takahashi, H., Carlson, K., Lagerback, P., Nystrom, A., Stevens, A., Rhoton, J., Orsini, G., Ouhammouch, M., Caer, J. Le, Brody, E., Adelman, K., Orsini, G., Kolb, A., Graziani, L., Brody, E., Severinova, E., Severinov, K., Fenyo, D., Marr, M., Brody, E., Roberts, J., Chait, B., Darst, S., Colland, F., Orsini, G., Brody, E., Buc, H., Kolb, A., Severinov, K., Muir, T.,

Sharma, U., Ravishankar, S., Shandil, R., Praveen, P., Balganesh, T., Urbauer, J., Adelman, K., Urbauer, R., Simeonov, M., Gilmore, J., Zolkiewski, M., Brody, E., Pahari, S., Chatterji, D., Simeonov, M., Urbauer, R.B., Gilmore, J., Adelman, K., Brody, E., Niedziela-Majka, A., Minakhin, L., Heyduk, T., Urbauer, J., Sharma, U., Chatterji, D., Minakhin, L., Niedziela-Majka, A., Kuznedelov, K., Adelman, K., Urbauer, J., Heyduk, T., Severinov, K., Dove, S., Hochschild, A., Urbauer, J., Simeonov, M., Urbauer, R., Adelman, K., Gilmore, J., Brody, E., Pal, D., Vuthoori, M., Pande, S., Wheeler, D., Hinton, D., Lambert, L., Schirf, V., Demeler, B., Cadene, M., Werner, M., Gilmore, J., Urbauer, R.B., Minakhin, L., Akoyev, V., Zolkiewski, M., Severinov, K., Urbauer, J., Sharma, U., Praveen, P., Balganesh, T., Stevens, A., Baxter, K., Lee, J., Minakhin, L., Severinov, K., Hinton, D., Minakhin, L., Camarero, J., Holford, M., Parker, C., Muir, T., Severinov, K., Gregory, B., Nickels, B., Garrity, S., Severinova, E., Minakhin, L., Urbauer, R., Urbauer, J., Heyduk, T., Severinov, K., Hochschild, A., Severinova, E., Severinov, K., Darst, S., Leirmo, S., Harrison, C., Cayley, D., Burgess, R., Record, M., Zou, L., Richardson, J., Orsini, G., Kolb, A., Buc, H., Orsini, G., Igonet, S., Pene, C., Sclavi, B., Buckle, M., Uzan, M., Kolb, A., Kolesky, S., Ouhammouch, M., Brody, E., Geiduschek, E., Mattson, T., Richardson, J., Goodin, D., Mattson, T., Houwe, G. Van, Epstein, R., Uzan, M., Brody, E., Favre, R., Hinton, D., March-Amegadzie, R., Gerber, J., Sharma, M., March-Amegadzie, R., Hinton, D., Stoskiene, G., Truncaite, L., Zajanckauskaite, A., Nivinskas, R., Cicero, M., Alexander, K., Kreuzer, K., Li, N., Sickmier, E., Zhang, R., Joachimiak, A., White, S., Li, N., Zhang, W., White, S., Kriwacki, R., Finnin, M., Hoffman, D., Kreuzer, K., Porter, S., Schmidt, R., White, S., Gerber, J., Hinton, D., Pande, S., Makela, A., Dove, S., Nickels, B., Hochschild, A., Hinton, D., Finnin, M., Cicero, M., Davies, C., Porter, S., White, S., Kreuzer, K., Guild, N., Gayle, M., Sweeney, R., Hollingsworth, T., Modeer, T., Gold, L., Marshall, P., Sharma, M., Hinton, D., Truncaite, L., Piesiniene, L., Kolesinskiene, G., Zajanckauskaite, A., Driukas, A., Klausa, V., Nivinskas, R., Truncaite, L., Zajanckauskaite, A., Nivinskas, R., Sharma, M., Marshall, P., Hinton, D., Bonocora, R., Caignan, G., Woodrell, C., Werner, M., Hinton, D., Kuldell, N., Hochschild, A., Li, M., Moyle, H., Susskind, M., Nickels, B., Dove, S., Murakami, K., Darst, S., Hochschild, A., Rhodius, V., Busby, S., Lonetto, M., Rhodius, V., Lamberg, K., Kiley, P., Busby, S., Gross, C., Landini, P., Busby, S.,

Bhende, P., Egan, S., Wickstrum, J., Egan, S., Decker, K., Hinton, D., Pulitzer, J., Coppo, A., Caruso, M., Cicero, M., Sharp, M., Gross, C., Kreuzer, K., Yuan, A., Nickels, B., Hochschild, A., Yuan, A., Hochschild, A., Yuan, A., Hochschild, A., Pineda, M., Gregory, B., Szczypinski, B., Baxter, K., Hochschild, A., Miller, E., Hinton, D., James, T., Cashel, M., Hinton, D., Adelman, K., Brody, E., Buckle, M., Hinton, D., Vuthoori, S., Wade, J., Struhl, K., Miller, E., Heidelberg, J., Eisen, J., Nelson, W., Durkin, A., Ciecko, A., Feldblyum, T., White, O., Paulsen, I., Nierman, W., Rao, X., Deighan, P., Hua, Z., Hu, X., Wang, J., Luo, M., Liang, Y., Zhong, G., Hochschild, A., Shen, L., Menkens, A., Kreuzer, K., Benson, K., Kreuzer, K., Luke, K., Radek, A., Liu, X., Campbell, J., Uzan, M., Haselkorn, R., Kogan, Y., Carpousis, A., Mudd, E., Krisch, H., Sanson, B., Uzan, M., Chiurazzi, M., Pulitzer, J., Dodd, I., Shearwin, K., Egan, J., Washburn, R., Stitt, B., Sozhamannan, S., Stitt, B., Glisovic, T., Bachorik, J., Yong, J., Dreyfuss, G., Zhang, Z., Klatt, A., Henderson, A., Gilmour, D., Ebright, R., Werner, F., Freiman, R., Albright, S., Zheng, S., Sha, W., Hammer, R., Tjian, R., Isogai, Y., Keles, S., Prestel, M., Hochheimer, A., Tjian, R., Daegelen, P., Brody, E., Daegelen, P., Brody, E., Hsu, T., Karam, J., Barth, K., Powell, D., Trupin, M., Mosig, G., Gruidl, M., Mosig, G., Nivinskas, R., Raudonikene, A., Guild, N., Gott, J., Zeeh, A., Bell-Pedersen, D., Ehrenman, K., Belfort, M., Shub, D., Tseng, M., He, P., Hilfinger, J., Greenberg, G., Belin, D., Mudd, E., Prentki, P., Yi-Yi, Y., Krisch, H., Estrem, S., Ross, W., Gaal, T., Chen, Z., Niu, W., Ebright, R. & Gourse, R. 2010. Transcriptional control in the prereplicative phase of T4 development. *Virology Journal*, 7(1): 289.

Hohn, B. & Murray, K. 1977. Packaging recombinant DNA molecules into bacteriophage particles in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, 74(8): 3259–3263.

Holmstrøm, K. & Rasmussen, O.F. 1990. An easy method to check the efficiency of biotin end-labelling of DNA-fragments. *Nucleic acids research*, 18(15): 4632.

Horgan, M., O'Flynn, G., Garry, J., Cooney, J., Coffey, A., Fitzgerald, G.F., Paul Ross, R. & McAuliffe, O. 2009. Phage lysin LysK can be truncated to its CHAP domain and retain lytic activity against live antibiotic-resistant staphylococci. *Applied and Environmental Microbiology*, 75(3): 872–874.

Horikoshi, K. 1999. Alkaliphiles: some applications of their products for biotechnology. *Microbiology and molecular biology reviews : MMBR*, 63(4): 735–750.

Hosoya, S., Adachi, K. & Kasai, H. 2009. Thalassomonas actiniarum sp. nov. and Thalassomonas haliotis sp. nov., isolated from marine animals. *International Journal of Systematic and Evolutionary Microbiology*, 59(4): 686–690.

Hou, S., Saw, J.H., Lee, K.S., Freitas, T. a, Belisle, C., Kawarabayasi, Y., Donachie, S.P., Pikina, A., Galperin, M.Y., Koonin, E. V, Makarova, K.S., Omelchenko, M. V, Sorokin, A., Wolf, Y.I., Li, Q.X., Keum, Y.S., Campbell, S., Denery, J., Aizawa, S.-I., Shibata, S., Malahoff, A. & Alam, M. 2004. Genome sequence of the deep-sea gamma-proteobacterium Idiomarina loihiensis reveals amino acid fermentation as a source of carbon and energy. *Proceedings of the National Academy of Sciences of the United States of America*, 101(52): 18036–18041.

Hsiao, C.L. & Black, L.W. 1978. Head morphogenesis of bacteriophage T4 III. The role of gene 20 in DNA packaging. *Virology*, 91(1): 26–38.

Hu, B., Margolin, W., Molineux, I.J. & Liu, J. 2015. Structural remodeling of bacteriophage T4 and host membranes during infection initiation. *Proceedings of the National Academy of Sciences of the United States of America*, 112(35): E4919–28.

Hua, N.P., Hamza-Chaffai, A., Vreeland, R.H., Isoda, H. & Naganuma, T. 2008. Virgibacillus salarius sp. nov., a halophilic bacterium isolated from a Saharan salt lake. *International Journal of Systematic and Evolutionary Microbiology*, 58(10): 2409–2414.

Huang, C.Y., Garcia, J.L., Patel, B.K.C., Cayol, J.L., Baresi, L. & Mah, R.A. 2000. Salinivibrio costicola subsp. vallismortis subsp. nov., a halotolerant facultative anaerobe from Death Valley, and emended description of Salinivibrio costicola. *International Journal of Systematic and Evolutionary Microbiology*, 50(2): 615–622.

Hugenholtz, P. 2002. Exploring prokaryotic diversity in the genomic era. *Genome biology*, 3(2): 3.

Hughes, K.A., Sutherland, I.W., Clark, J. & Jones, M. V. 1998. Bacteriophage and associated

polysaccharide depolymerases - Novel tools for study of bacterial biofilms. *Journal of Applied Microbiology*, 85(3): 583–590.

Hurwitz, B.L., Deng, L., Poulos, B.T. & Sullivan, M.B. 2013. Evaluation of methods to concentrate and purify ocean virus communities through comparative, replicated metagenomics. *Environmental Microbiology*, 15(5): 1428–1440.

Hutchinson, P. 1995. The climatology of Namibia and its relevance to the drought situation. *Coping with aridity: Drought impacts and preparedness in Namibia*: 17–38.

Hutchison, C. a, Smith, H.O., Pfannkoch, C. & Venter, J.C. 2005. Cell-free cloning using phi29 DNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America*, 102(48): 17332–17336.

Hyman, P. & Abedon, S.T. 2009. *Bacteriophages*. M. R. J. Clokie & A. M. Kropinski, eds. Totowa, NJ: Humana Press.

Inouye, S. & Nakamura, M. 2003. Identification of biotinylated lysine residues in the photoprotein aequorin by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry peptide mapping after lysine-specific endopeptidase digestion. *Analytical Biochemistry*, 316(2): 216–222.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakatura, A. 1987. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isoenzyme conversion in Escherichia coli, and identification of the gene product. , 169(12): 5429–5433.

Jacobs, W., Barletta, R., Udani, R., Chan, J., Kalkut, G., Sosne, G., Kieser, T., Sarkis, G., Hatfull, G. & Bloom, B. 1993. Rapid assessment of drug susceptibilities of Mycobacterium tuberculosis by means of luciferase reporter phages. *Science*, 260(5109): 819–822.

Jacobson, A. 1972. Role of F pili in the penetration of bacteriophage fl. *Journal of virology*, 10(4): 835–43.

Jansen, R., Van Embden, J.D.A., Gaastra, W. & Schouls, L.M. 2002. Identification of a novel family of sequence repeats among prokaryotes. , 6(1): 23–33.

Jiang, S.C. & Paul, J.H. 1994. Seasonal and diel abundance of viruses and occurrence of lysogeny/bacteriocinogeny in the marine environment. *Marine Ecology Progress Series*, 104(1-2): 163–172.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. & Charpentier, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. , 337(6096): 816–821.

Jones, D.T., Taylor, W.R. & Thornton, J.M. 1992. The rapid generation of mutation data matrices from protein sequences. *Computer applications in the biosciences : CABIOS*, 8(3): 275–82.

Kalinski, A. & Black, L.W. 1986. End structure and mechanisms of packaging of bacteriophage T4 DNA. , 58(3): 951–954.

Kao, S.H. & McClain, W.H. 1980. Roles of bacteriophage T4 gene 5 and gene s products in cell lysis. *Journal of virology*, 34(1): 104–7.

Kapfhammer, D., Blass, J., Evers, S. & Reidl, J. 2002. Vibrio cholerae phage K139: complete genome sequence and comparative genomics of related phages. *Journal of bacteriology*, 184(23): 6592–601.

Kasprzak, J.M., Czerwoniec, A., Bujnicki, J.M., Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A., Steinberg, S., Kowalak, J., Bruenger, E., McCloskey, J., Dalluge, J., Hashizume, T., Sopchik, A., McCloskey, J., Davis, D., Dalluge, J., Hamamoto, T., Horikoshi, K., Morita, R., Stetter, K., McCloskey, J., Kato, T., Daigo, Y., Hayama, S., Ishikawa, N., Yamabuki, T., Ito, T., Miyamoto, M., Kondo, S., Nakamura, Y., Bishop, A., Xu, J., Johnson, R., Schimmel, P., Crecy-Lagard, V. de, Czerwoniec, A., Dunin-Horkawicz, S., Purta, E., Kaminska, K., Kasprzak, J., Bujnicki, J., Grosjean, H., Rother, K., Xing, F., Hiley, S., Hughes, T., Phizicky, E., Xing, F., Martzen, M., Phizicky, E., Savage, D., Crecy-Lagard, V. de, Bishop, A., Dobritzsch, D., Ricagno, S., Schneider, G., Schnackerz, K., Lindqvist, Y., Rowland, P., Bjornberg, O., Nielsen, F., Jensen, K., Larsen, S., Yu, F., Tanaka, Y., Yamashita, K., Suzuki, T., Nakamura, A., Hirano, N., Suzuki, T., Yao, M., Tanaka, I., Altschul, S., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W., Lipman, D., Frickey, T., Lupas, A., Park, F., Gajiwala, K., Noland, B., Wu, L., He, D., Molinari, J., Loomis,

K., Pagarigan, B., Kearins, P., Christopher, J., Thompson, J., Gibson, T., Plewniak, F., Jeanmougin, F., Higgins, D., Kurowski, M., Bujnicki, J., Tatusov, R., Fedorova, N., Jackson, J., Jacobs, A., Kiryutin, B., Koonin, E., Krylov, D., Mazumder, R., Mekhedov, S., Nikolskaya, A., Ronquist, F., Huelsenbeck, J., Shatkay, H., Hoglund, A., Brady, S., Blum, T., Donnes, P., Kohlbacher, O., Horton, P., Park, K., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C., Nakai, K., Chou, K., Shen, H., Hawkins, J., Davis, L., Boden, M., Emanuelsson, O., Brunak, S., Heijne, G. von, Nielsen, H., Yu, C., Lin, C., Hwang, J., Glaser, F., Pupko, T., Paz, I., Bell, R., Bechor-Shental, D., Martz, E., Ben-Tal, N., Baker, N., Sept, D., Joseph, S., Holst, M., McCammon, J., DeLano, W., Rider, L., Ottosen, M., Gattis, S., Palfey, B., Porter, D., Chestnut, W., Taylor, L., Merrill, B., Spector, T., Tamura, K., Dudley, J., Nei, M., Kumar, S., Whelan, S., Goldman, N., Lundstrom, J., Rychlewski, L., Bujnicki, J., Elofsson, A., Wallner, B., Elofsson, A., Sali, A., Blundell, T., Pawlowski, M., Gajda, M., Matlak, R., Bujnicki, J., Misura, K., Chivian, D., Rohl, C., Kim, D., Baker, D., Wallner, B., Fang, H., Elofsson, A., Kozlowski, L., Bujnicki, J., Linding, R., Jensen, L., Diella, F., Bork, P., Gibson, T., Russell, R., Dunker, A., Brown, C., Lawson, J., Iakoucheva, L., Obradovic, Z., Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., Dunker, A., Dosztanyi, Z., Csizmok, V., Tompa, P., Simon, I., Shimizu, K., Hirose, S., Noguchi, T., Hirose, S., Shimizu, K., Kanai, S., Kuroda, Y., Noguchi, T., Ishida, T., Kinoshita, K., Yang, Z., Thomson, R., McNeil, P., Esnouf, R., Puton, T., Kozlowski, L., Tuszynska, I., Rother, K., Bujnicki, J., Murakami, Y., Spriggs, R., Nakamura, H., Jones, S., Kumar, M., Wang, L., Huang, C., Yang, M. & Yang, J. 2012. Molecular evolution of dihydrouridine synthases. *BMC Bioinformatics*, 13(1): 153.

Katsura, I. & Hendrix, R.W. 1984. Length determination in bacteriophage lambda tails. *Cell*, 39(3 PART 2): 691–698.

Katsura, I. & Kobayashi, H. 1990. Structure and inherent properties of the bacteriophage lambda head shell. VII. Molecular design of the form-determining major capsid protein. *Journal of Molecular Biology*, 213(3): 503–511.

Kellenberger, E., Stauffer, E., Häner, M., Lustig, A. & Karamata, D. 1996. Mechanism of the long tail-

fiber deployment of bacteriophages T-even and its role in adsorption, infection and sedimentation. *Biophysical Chemistry*, 59(1): 41–59.

Kim, M.S., Kim, Y.D., Hong, S.S., Park, K., Ko, K.S. & Myung, H. 2015. Phage-Encoded Colanic Acid-Degrading Enzyme Permits Lytic Phage Infection of a Capsule-Forming Resistant Mutant Escherichia coli Strain K. E. Wommack, ed. *Applied and Environmental Microbiology*, 81(3): 900–909.

Kirby, J.E., Trempy, J.E. & Gotfiesmanl, S. 1994. Expression in Escherichia coli. *Journal of Bacteriology*, (301): 2068–2081.

Klement, J.F., Moorefiedl, M.B., Jorgensen, E., Brown, J.E., Risman, S. & McAllister, W.T. 1990. Discrimination between bacteriophage T3 and T7 promoters by the T3 and T7 RNA polymerases depends primarily upon a three base-pair region located 10 to 12 base-pairs upstream from the start site. *Journal of Molecular Biology*, 215(1): 21–29.

Knipe, D. & Howley, P. 2001. *Fundamental virology*. 4th ed. Philidelphia: Lippincott Williams & Wilkins.

Koide, A., Wojcik, J., Gilbreth, R.N., Reichel, A., Piehler, J. & Koide, S. 2009. Accelerating phage-display library selection by reversible and site-specific biotinylation. *Protein Engineering, Design and Selection*, 22(11): 685–690.

Korlach, J., Bibillo, A., Wegener, J., Peluso, P., Pham, T.T., Park, I., Clark, S., Otto, G. a & Turner, S.W. 2008. Long, processive enzymatic DNA synthesis using 100% dye-labeled terminal phosphate-linked nucleotides. *Nucleosides, nucleotides & nucleic acids*, 27(9): 1072–1083.

Kostyuchenko, V.A., Chipman, P.R., Leiman, P.G., Arisaka, F., Mesyanzhinov, V. V & Rossmann, M.G. 2005. The tail structure of bacteriophage T4 and its mechanism of contraction. *Nature Structural & Molecular Biology*, 12(9): 810–813.

Kreuzer, K.N. 2000. Recombination-dependent DNA replication in phage T4. *Trends in Biochemical Sciences*, 25(4): 165–173.

Kreuzer, K.N. & Brister, J.R. 2010. Initiation of bacteriophage T4 DNA replication and replication fork dynamics: a review in the Virology Journal series on bacteriophage T4 and its relatives. *Virology journal*, 7: 358.

Krienitz, L. & Kotut, K. 2010. Fluctuating algal food populations and the occurrence of lesser flamingos (phoeniconaias minor) in three Kenyan rift valley lakes. *Journal of Phycology*, 46(6): 1088–1096.

Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E.L.L. 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of molecular biology*, 305(3): 567–580.

Kropinski, A.M. 2009. Bacteriophages. *Methods in molecular biology*, 502: 365–370.

Krüger, D.H. & Bickle, T.A. 1983. Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiological reviews*, 47(3): 345–360.

Kutter, E., Kellenberger, E., Carlson, K., Eddy, S., Neitzel, J., Messinge, L., North, J. & Guttman, B. 1994. Effects of bacterial growth condition and physiology on T4 infection. In J. Karam, ed. *Molecular Biology of Bacteriophage T4*. Washington ,D.C.: American Society for Microbiology,: 406–418.

Kwoh, D.Y. & Kemper, J. 1978. Bacteriophage P22-mediated specialized transduction in Salmonella typhimurium: high frequency of aberrant prophage excision. *Journal of virology*, 27(3): 519–34.

Labonté, J.M., Reid, K.E. & Suttle, C.A. 2009. Phylogenetic analysis indicates evolutionary diversity and environmental segregation of marine podovirus DNA polymerase gene sequences. *Applied and Environmental Microbiology*, 75(11): 3634–3640.

Labonté, J.M., Swan, B.K., Poulos, B., Luo, H., Koren, S., Hallam, S.J., Sullivan, M.B., Woyke, T., Eric Wommack, K. & Stepanauskas, R. 2015. Single-cell genomics-based analysis of virus–host interactions in marine surface bacterioplankton. *The International Society for Microbial Ecology Journal*, 9(11): 2386–2399.

Laity, J.H., Lee, B.M. & Wright, P.E. 2001. Zinc finger proteins: New insights into structural and

functional diversity. *Current Opinion in Structural Biology*, 11(1): 39–46.

Lancaster, J., Lancaster, N. & Seely, M.K. 1983. Climate of the central Namib Desert M. Seely, ed. *Deserts—The Illustrated Library of the Earth*: 14– 25.

Landthaler, M. & Shub, D.A. 1999. Unexpected abundance of self-splicing introns in the genome of bacteriophage Twort: introns in multiple genes, a single gene with three introns, and exon skipping by group I ribozymes. *Proceedings of the National Academy of Sciences of the United States of America*, 96(12): 7005–10.

Lang, A.S., Zhaxybayeva, O. & Beatty, J.T. 2012. Gene transfer agents: phage-like elements of genetic exchange. *Nature Reviews Microbiology*, 10(7): 472–82.

Langlois, C., Ramboarina, S., Cukkemane, A., Auzat, I., Chagot, B., Gilquin, B., Ignatiou, A., Petitpas, I., Kasotakis, E., Paternostre, M., White, H.E., Orlova, E. V, Baldus, M., Tavares, P. & Zinn-Justin, S. 2015. Bacteriophage SPP1 tail tube protein self-assembles into β-structure-rich tubes. *The Journal of biological chemistry*, 290(6): 3836–49.

Larocca, D., Burg, M.A., Jensen-Pergakes, K., Ravey, E.P., Gonzalez, A.M. & Baird, A. 2002. Evolving phage vectors for cell targeted gene delivery. *Current pharmaceutical biotechnology*, 3(1): 45–57.

Laslett, D. & Canback, B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research*, 32(1): 11–16.

Ledford, H. 2016. Riding the CRISPR wave. *Nature*, 531: 156.

Ledford, H. 2015. Where in the world could the first CRISPR baby be born? *Nature*, 526(7573): 310–311.

Lee, C.N., Tseng, T.-T., Chang, H.C., Lin, J.W. & Weng, S.F. 2014. Genomic sequence of temperate phage Smp131 of Stenotrophomonas maltophilia that has similar prophages in xanthomonads. *BMC microbiology*, 14(1): 17.

Lee, L.H., Lui, D., Platner, P.J., Hsu, S.-F., Chu, T.-C., Gaynor, J.J., Vega, Q.C. & Lustigman, B.K. 2006. Induction of temperate cyanophage AS-1 by heavy metal--copper. *BMC microbiology*, 6(1): 17.

Lee, O.O., Lai, P.Y., Wu, H. xian, Zhou, X. jian, Miao, L., Wang, H. & Qian, P.Y. 2012. Marinobacter xestospongiae sp. nov., isolated from the marine sponge Xestospongia testudinaria collected from the Red sea. *International Journal of Systematic and Evolutionary Microbiology*, 62(8): 1980–1985.

Leiman, P., Arisaka, F. & Raaij, M. van. 2010. Morphogenesis of the T4 tail and tail fibers. *Virology*.

Leiman, P.G., Chipman, P.R., Kostyuchenko, V.A., Mesyanzhinov, V. V. & Rossmann, M.G. 2004. Three-Dimensional Rearrangement of Proteins in the Tail of Bacteriophage T4 on Infection of Its Host. *Cell*, 118(4): 419–429.

Li, J. & Dennehy, J.J. 2011. Differential bacteriophage mortality on exposure to copper. *Applied and environmental microbiology*, 77(19): 6878–83.

Lieu, P.T., Jozsi, P., Gilles, P. & Peterson, T. 2005. Development of a DNA-labeling system for array-based comparative genomic hybridization. *Journal of biomolecular techniques : JBT*, 16(2): 104–11.

Lima-Mendez, G., Toussaint, A. & Leplae, R. 2011. A modular view of the bacteriophage genomic space: identification of host and lifestyle marker modules. *Research in microbiology*, 162(8): 737–46.

Lima-Mendez, G., Toussaint, A. & Leplae, R. 2007. Analysis of the phage sequence space: The benefit of structured information. *Virology*, 365(2): 241–249.

Lindahl, G. & Sunshine, M. 1972. Excision-deficient mutants of bacteriophage P2. *Virology*, 49(1): 180–187.

Linderoth, N.A., Ziermann, R., Haggård-Ljungquist, E., Christie, G.E. & Calendar, R. 1991. Nucleotide sequence of the DNA packaging and capsid synthesis genes of bacteriophage P2. *Nucleic Acids*

Research, 19(25): 7207–7214.

Liu, C.C. & Alberts, B.M. 1980. Pentaribonucleotides of mixed sequence are synthesized and efficiently prime de novo DNA chain starts in the T4 bacteriophage DNA replication system. *Proceedings of the National Academy of Sciences of the United States of America*, 77(10): 5698–702.

Liu, J., Glazko, G. & Mushegian, A. 2006. Protein repertoire of double-stranded DNA bacteriophages. *Virus Research*, 117(1): 68–80.

Liu, R., Zhang, P., Pu, X., Xing, X. & Pathology, P. 2011. Analysis of a Prophage Gene Frequency Revealed Population Variation of ' *Candidatus* Liberibacter asiaticus ' from Two Citrus-Growing Provinces in China. *Plant Disease*, 95(November 2010): 431–435.

Lobry, J.R. 1996. Asymmetric substitution patterns in the two DNA strands of bacteria. *Molecular biology and evolution*, 13(5): 660–5.

Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D. & Darnell, J. 2000. Constructing DNA Libraries with λ Phage and Other Cloning Vectors. http://www.ncbi.nlm.nih.gov/books/NBK21696/ 14 July 2016.

Łoś, M. & Węgrzyn, G. 2012. Pseudolysogeny. *Advances in virus research*, 82: 339–49.

Los, M., Wegrzyn, G. & Neubauer, P. 2003. A role for bacteriophage T4 rI gene function in the control of phage development during pseudolysogeny and in slowly growing host cells. *Research in microbiology*, 154(8): 547–52.

Lowe, T.M. & Eddy, S.R. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, 25(5): 955–964.

Łyzeń, R., Kochanowska, M., Węgrzyn, G. & Szalewska-Pałasz, A. 2009. Transcription from bacteriophage λ pR promoter is regulated independently and antagonistically by DksA and ppGpp. *Nucleic Acids Research*, 37(20): 6655–6664.

Macechko, P.T., Krueger, L., Hirsch, B. & Erlandsen, S.L. 1997. Comparison of Immunologic

Amplification vs Enzymatic Deposition of Fluorochrome-conjugated Tyramide as Detection Systems for FISH. *Journal of Histochemistry & Cytochemistry*, 45(3): 359–363.

Machado, A., Almeida, C., Carvalho, A., Boyen, F., Haesebrouck, F., Rodrigues, L., Cerca, N. & Azevedo, N.F. 2013. Fluorescence in situ hybridization method using a peptide nucleic acid probe for identification of Lactobacillus spp. in milk samples. *International Journal of Food Microbiology*, 162(1): 64–70.

Macián, M.C., Ludwig, W., Schleifer, K.H., Garay, E. & Pujalte, M.J. 2001. Thalassomonas viridans gen. nov., sp. nov., a novel marine gamma-proteobacterium. *International journal of systematic and evolutionary microbiology*, 51(Pt 4): 1283–1289.

Makhalanyane, T.P. 2012. *Microbial ecology of hot and cold desert edaphic communities*. University of the Western Cape, Cape Town, South Africa.

Makhalanyane, T.P., Valverde, A., Lacap, D.C., Pointing, S.B., Tuffin, M.I. & Cowan, D.A. 2013. Evidence of species recruitment and development of hot desert hypolithic communities. *Environmental Microbiology Reports*, 5(2): 219–224.

Malki, K., Kula, A., Bruder, K., Sible, E., Hatzopoulos, T., Steidel, S., Watkins, S.C. & Putonti, C. 2015. Bacteriophages isolated from Lake Michigan demonstrate broad host-range across several bacterial phyla. *Virology journal*, 12(1): 164.

Mann, N.H., Clokie, M.R.J., Millard, A., Cook, A., Wilson, W.H., Wheatley, P.J., Letarov, A. & Krisch, H.M. 2005. The genome of S-PM2, a 'photosynthetic' T4-type bacteriophage that infects marine Synechococcus strains. *Journal of Bacteriology*, 187(9): 3188–3200.

Manti, A., Boi, P., Amalfitano, S., Puddu, A. & Papa, S. 2011. Experimental improvements in combining CARD-FISH and flow cytometry for bacterial cell quantification. *Journal of Microbiological Methods*, 87(3): 309–315.

Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., Lanczycki, C.J., Lu, F., Marchler, G.H., Song, J.S., Thanki, N.,

Wang, Z., Yamashita, R.A., Zhang, D., Zheng, C. & Bryant, S.H. 2015. CDD: NCBI's conserved domain database. *Nucleic Acids Research*, 43(D1): D222–D226.

Marinus, M.G. & Morris, N.R. 1973. Isolation of deoxyribonucleic acid methylase mutants of Escherichia coli K-12. *Journal of bacteriology*, 114(3): 1143–50.

Martínez-García, M., Santos, F., Moreno-Paz, M., Parro, V. & Antón, J. 2014. Unveiling viral-host interactions within the 'microbial dark matter'. *Nature communications*, 5: 4542.

Martinsohn, J.T., Radman, M. & Petit, M.A. 2008. The ?? red proteins promote efficient recombination between diverged sequences: Implications for bacteriophage genome mosaicism. *PLoS Genetics*, 4(5): e1000065.

May, M.S. & Hattman, S. 1975. Analysis of bacteriophage deoxyribonucleic acid sequences methylated by host- and R-factor-controlled enzymes. *Journal of bacteriology*, 123(2): 768–70.

Mayer, G. 2007. Microbiology and Immunology. *Microbiology and Immunology Online, School of Medicine University of South Carolina*. http://microbiologybook.org/book/immunol-sta.htm 5 November 2014.

McEntee, K. 1976. Specialized transduction of recA by bacteriophage lambda. *Virology*, 70(1): 221–222.

McNamara, A.R. 2000. A novel four zinc-finger protein targeted against p190BcrAbl fusion oncogene cDNA: utilisation of zinc-finger recognition codes. *Nucleic Acids Research*, 28(24): 4865–4872.

Menkens, A.E. & Kreuzer, K.N. 1988. Deletion analysis of bacteriophage T4 tertiary origins. A promoter sequence is required for a rifampicin-resistant replication origin. *The Journal of biological chemistry*, 263(23): 11358–65.

Middelboe, M. 2008. Ch. 11 -- Microbial Disease in the Sea: Effects of Viruses on Carbon and Nutrient Cycling. *Infectious disease ecology, effects of ecosystems on desease and of disease on ecosystems*: 242–259.

Miller, E.S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T. & Ru, W. 2003. Bacteriophage T4 Genome †. *Microbiology and Molecular Biology Reviews*, 67(1): 86–156.

Miller, R. & Day, M. 2008. in. In S. T. Abedon, ed. *Bacteriophage Ecology*. Cambridge Univ. Press: 114–144.

Mizuno, C.M., Rodriguez-Valera, F., Kimes, N.E. & Ghai, R. 2013. Expanding the Marine Virosphere Using Metagenomics. *PLoS Genetics*, 9(12).

Mizuuchi, K., Kemper, B., Hays, J. & Weisberg, R.A. 1982. T4 endonuclease VII cleaves holliday structures. , 29(2): 357–365.

Mojica, F.J.M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. , 60(2): 174–182.

Mokrousov, I. 2009. Corynebacterium diphtheriae: Genome diversity, population structure and genotyping perspectives. *Infection, Genetics and Evolution*, 9(1): 1–15.

Moldovan, R., Chapman-McQuiston, E. & Wu, X.L. 2007. On kinetics of phage adsorption. *Biophysical journal*, 93(1): 303–315.

Molineux, I.J. 2001. No syringes please, ejection of phage T7 DNA from the virion is enzyme driven. *Molecular microbiology*, 40(1): 1–8.

Mondragon, A. & Harrison, S.C. 1991. The Phage 434 Cro/OR1 Complex at 2.35 Å Resolution. *J. Molec. Biol.*, 219(3): 321–334.

Montag, D., Hashemolhosseini, S. & Henning, U. 1990. Receptor-recognizing proteins of T-even type bacteriophages: The receptor-recognizing area of proteins 37 of phages T4 TuIa and TuIb. *Journal of Molecular Biology*, 216(2): 327–334.

Moraru, C., Lam, P., Fuchs, B.M., Kuypers, M.M.M. & Amann, R. 2010. GeneFISH - an in situ technique for linking gene presence and cell identity in environmental microorganisms. *Environmental Microbiology*, 12(11): 3057–3073.

Morris, P., Marinelli, L.J., Jacobs-Sera, D., Hendrix, R.W. & Hatfull, G.F. 2008. Genomic characterization of mycobacteriophage giles: Evidence for phage acquisition of host DNA by illegitimate recombination. *Journal of Bacteriology*, 190(6): 2172–2182.

Moser, M.J., DiFrancesco, R.A., Gowda, K., Klingele, A.J., Sugar, D.R., Stocki, S., Mead, D.A. & Schoenfeld, T.W. 2012. Thermostable DNA polymerase from a viral metagenome is a potent RT-PCR enzyme. *PLoS ONE*, 7(6): e38371.

Mosier-Boss, P.A., Lieberman, S.H., Andrews, J.M., Rohwer, F.L., Wegley, L.E. & Breitbart, M. 2003. Use of Fluorescently Labeled Phage in the Detection and Identification of Bacterial Species. *Applied Spectroscopy*, 57(9): 1138–1144.

Mosig, G., Lin, G., Franklin, J. & Fan, W. 1989. Functional relationships and structural determinants of two bacteriophage T4 lysozymes: a soluble (gene e) and a baseplate-associated (gene 5) protein. *The New biologist*.

Moter, A. & Göbel, U.B. 2000. Fluorescence in situ hybridization (FISH) for direct visualization of microorganisms. *Journal of Microbiological Methods*, 41(2): 85–112.

Murialdo, H. 1977. Assembly of Biologically Active Proheads of Bacteriophage Lambda in vitro. *Proceedings of the National Academy of Sciences*, 74(3): 906–910.

Mwirichia, R., Muigai, A.W., Tindall, B., Boga, H.I. & Stackebrandt, E. 2010. Isolation and characterisation of bacteria from the haloalkaline Lake Elmenteita, Kenya. *Extremophiles*, 14(4): 339–348.

Necşulea, A. & Lobry, J.R. 2007. A new method for assessing the effect of replication on DNA base composition asymmetry. *Molecular Biology and Evolution*, 24(10): 2169–2179.

Nelson, D.C., Schmelcher, M., Rodriguez-Rubio, L., Klumpp, J., Pritchard, D.G., Dong, S. & Donovan, D.M. 2012. Endolysins as Antimicrobials. *Advances in Virus Research*, 83: 299–365.

Nelson, E.J. & Ghiorse, W.C. 2002. Isolation and identification of Pseudoalteromonas piscicida strain Cura-d associated with diseased damselfish (Pomacentridae) eggs. *Journal of Fish Diseases*,

22(4): 253–260.

Nisole, S. & Saïb, A. 2004. Early steps of retrovirus replicative cycle. *Retrovirology*, 1(1): 9.

Nossal, N.G. 1980. RNA priming of DNA replication by bacteriophage T4 proteins. *The Journal of biological chemistry*, 255(5): 2176–82.

O'Flaherty, S., Coffey, A., Meaney, W., Fitzgerald, G.F. & Ross, R.P. 2005. The recombinant phage lysin LysK has a broad spectrum of lytic activity against clinically relevant staphylococci, including methicillin-resistant Staphylococcus aureus. *Journal of Bacteriology*, 187(20): 7161–7164.

O'Keefe, C.L., Warburton, P.E. & Gregory Matera, A. 1996. Oligonucleotide probes for alpha satellite DNA variants can distinguish homologous chromosomes by FISH. *Human Molecular Genetics*, 5(11): 1793–1799.

Ohno, S., Okano, H., Tanji, Y., Ohashi, A., Watanabe, K., Takai, K. & Imachi, H. 2012. A method for evaluating the host range of bacteriophages using phages fluorescently labeled with 5-ethynyl-2'-deoxyuridine (EdU). *Applied Microbiology and Biotechnology*, 95(3): 777–788.

Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K. & Sugino, A. 1968. Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Pnas*, 59(2): 598–605.

Oppenheim, A.B., Kobiler, O., Stavans, J., Court, D.L. & Adhya, S. 2005. Switches in bacteriophage lambda development. *Annual review of genetics*, 39: 409–429.

Østergaard, S., Brøndsted, L. & Vogensen, F.K. 2001. Identification of a replication protein and repeats essential for DNA replication of the temperate lactococcal bacteriophage TP901-1. *Applied and Environmental Microbiology*, 67(2): 774–781.

Overbaugh, J., Miller, A.D. & Eiden, M. V. 2001. Receptors and entry cofactors for retroviruses include single and multiple transmembrane-spanning proteins as well as newly described glycophosphatidylinositol-anchored and secreted proteins. *Microbiology and molecular biology*

reviews : MMBR, 65(3): 371–89.

Pan, W., Byrne-Steele, M., Wang, C., Lu, S., Clemmons, S., Zahorchak, R.J. & Han, J. 2014. DNA polymerase preference determines PCR priming efficiency. *BMC biotechnology*, 14(1): 10.

Park, S., Choi, W.C., Oh, T.K. & Yoon, J.H. 2011. Thalassomonas agariperforans sp. nov., an agarolytic bacterium isolated from marine sand. *International Journal of Systematic and Evolutionary Microbiology*, 61(11): 2573–2576.

Parua, P.K., Mondal, A. & Parrack, P. 2010. HflD, an Escherichia coli protein involved in the lambda lysis-lysogeny switch, impairs transcription activation by lambdaCII. *Archives of biochemistry and biophysics*, 493(2): 175–83.

Paul, J.H. 1999. Microbial gene transfer: an ecological perspective. *Journal of molecular microbiology and biotechnology*, 1(1): 45–50.

Paul, J.H. & Sullivan, M.B. 2005. Marine phage genomics: What have we learned? *Current Opinion in Biotechnology*, 16(3 SPEC. ISS.): 299–307.

Paul, J.H., Sullivan, M.B., Segall, A.M. & Rohwer, F. 2002. Marine phage genomics. *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*, 133(4): 463–476.

Paul, V.D., Rajagopalan, S.S., Sundarrajan, S., George, S.E., Asrani, J.Y., Pillai, R., Chikkamadaiah, R., Durgaiah, M., Sriram, B., Padmanabhan, S., V., P., S., R., S., S., S.E., G., J.Y., A., R., P., R., C., M., D., B., S., S., P., Paul V., Rajagopalan S., Sundarrajan S., George S.E., Asrani J.Y., Pillai R., Chikkamadaiah R., Durgaiah M., Sriram B. & Padmanabhan S. 2011. A novel bacteriophage Tail-Associated Muralytic Enzyme (TAME) from Phage K and its development into a potent antistaphylococcal protein. *BMC microbiology*, 11(1): 226.

Pedulla, M.L., Ford, M.E., Houtz, J.M., Karthikeyan, T., Wadsworth, C., Lewis, J.A., Jacobs-Sera, D., Falbo, J., Gross, J., Pannunzio, N.R., Brucker, W., Kumar, V., Kandasamy, J., Keenan, L., Bardarov, S., Kriakov, J., Lawrence, J.G., Jacobs, W.R., Hendrix, R.W. & Hatfull, G.F. 2003. Origins of highly mosaic mycobacteriophage genomes. *Cell*, 113(2): 171–182.

Peduzzi, P., Gruber, M., Gruber, M. & Schagerl, M. 2014. The virus's tooth: cyanophages affect an African flamingo population in a bottom-up cascade. *The ISME journal*, 8(6): 1346–51.

Pernthaler, A., Pernthaler, J. & Amann, R. 2002. Fluorescence In Situ Hybridization and Catalyzed Reporter Deposition for the Identification of Marine Bacteria Fluorescence In Situ Hybridization and Catalyzed Reporter Deposition for the Identification of Marine Bacteria. *Applied and Environmental Microbiology*, 68(6): 3094–3101.

Petersen, T.W. & van den Engh, G. 2003. Stability of the breakoff point in a high-speed cell sorter. *Cytometry. Part A : the journal of the International Society for Analytical Cytology*, 56(2): 63–70.

Pogliano, J., Ho, T.Q., Zhong, Z. & Helinski, D.R. 2001. Multicopy plasmids are clustered and localized in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 98(8): 4486–4491.

Poranen, M.M., Daugelavicius, R. & Bamford, D.H. 2002. Common principles in viral entry. *Annual review of microbiology*, 56: 521–38.

Pourcel, C., Salvignol, G. & Vergnaud, G. 2005. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. , 151(3): 653–663.

Powell, B.S., Rivas, M.P., Court, D.L., Nakamura, Y. & Turnbough, C.L. 1995. Rapid confirmation of single copy lambda prophage integration by PCR. *Nucleic Acids Research*, 23(7): 1278.

Prangishvili, D. & Krupovic, M. 2012. A new proposed taxon for double-stranded DNA viruses, the order 'Ligamenvirales'. *Archives of Virology*, 157(4): 791–795.

Prehm, P., Jann, B., Jann, K., Schmidt, G. & Stirm, S. 1976. On a bacteriophage T3 and T4 receptor region within the cell wall lipopolysaccharide of escherichia coli B. *Journal of Molecular Biology*, 101(2): 277–281.

Prescott, L. 1993. *Microbiology*. Wm. C. Brown Publishers.

Prestel, E., Regeard, C., Andrews, J., Oger, P. & DuBow, M.S. 2012. A Novel Bacteriophage Morphotype with a Ribbon-like Structure at the Tail Extremity. *Research Journal of Microbiology*, 7(1): 75–81.

Prestel, E., Regeard, C., Salamitou, S., Neveu, J. & Dubow, M.S. 2013. The bacteria and bacteriophages from a Mesquite Flats site of the Death Valley desert. *Antonie van Leeuwenhoek*, 103(6): 1329–41.

Prestel, E., Salamitou, S. & Dubow, M.S. 2008. An examination of the bacteriophages and bacteria of the Namib desert. *Journal of Microbiology*, 46(4): 364–372.

Pride, D.T., Wassenaar, T.M., Ghose, C. & Blaser, M.J. 2006. Evidence of host-virus co-evolution in tetranucleotide usage patterns of bacteriophages and eukaryotic viruses. *BMC genomics*, 7(1): 8.

Proctor, L.M. & Fuhrman, J.A. 1991. Roles of viral infection in organic particle flux. *Marine Ecology Progress Series*, 69(1-2): 133–142.

Pryshliak, M., Hammerl, J.A., Reetz, J., Strauch, E. & Hertwig, S. 2014. Vibrio vulnificus phage PV94 Is closely related to temperate phages of V. cholerae and other vibrio species. *PLoS ONE*, 9(4): e94707.

Ptashne, M., Jeffrey, A., Johnson, A., Maurer, R., Meyer, B., Pabo, C., Roberts, T. & Sauer, R. 1980. How the Lambda Repressor and Cro Work. *Cell*, 19: 1–11.

Rabbani, I., Radoja, U., Ray, S., Selim, A. & Weston, C. 2004. The Dynamics of Bacteriophage T4 Binding to Escherichia coli. *Journal of Experimental Microbiology and Immunology*, 5(April): 61–64.

Rajagopala, S. V, Casjens, S. & Uetz, P. 2011. The protein interaction map of bacteriophage lambda. *BMC Microbiology*, 11(1): 213.

Rakhuba, D. V., Kolomiets, E.I., Szwajcer Dey, E. & Novik, G.I. 2010. Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell. *Polish Journal of Microbiology*, 59(3): 145–155.

Ramond, J.-B., Pienaar, A., Armstrong, A., Seely, M. & Cowan, D.A. 2014. Niche-partitioning of edaphic microbial communities in the Namib Desert gravel plain Fairy Circles. *PloS one*, 9(10): e109539.

Rao, V. & Black, L. 2005. DNA packaging in bacteriophage T4. *Viral genome packaging machines: genettics, structure, and mechnism*: 40–58.

Rao, V.B. & Feiss, M. 2015. Mechanisms of DNA Packaging by Large Double-Stranded DNA Viruses. *Annual Review of Virology*, 2(1): 351–378.

Rath, D., Amlinger, L., Rath, A. & Lundgren, M. 2015. The CRISPR-Cas immune system: Biology, mechanisms and applications. *Biochimie*, 117: 119–128.

Reardon, S. 2014. Phage therapy gets revitalized. *Nature*, 510: 15–16.

Redon, R., Fitzgerald, T. & Carter, N.P. 2009. Comparative genomic hybridization: DNA labeling, hybridization and detection. *Methods in molecular biology (Clifton, N.J.)*, 529: 267–78.

Reidl, J. & Mekalanos, J.J. 1995. Characterization of Vibrio cholerae bacteriophage K139 and use of a novel mini-transposon to identify a phage-encoded virulence factor. *Molecular microbiology*, 18(4): 685–701.

Reysenbach, A.-L. & Pace, N.R. 1995. *Reliable amplification of hyperthermophilic Archaeal 16S rRNA genes by the polymerase chain reaction.*

Riechmann, L. & Holliger, P. 1997. The C-terminal domain of TolA is the coreceptor for filamentous phage infection of E. coli. *Cell*, 90(2): 351–360.

Riede, I., Degen, M. & Henning, U. 1985. The receptor specificity of bacteriophages can be determined by a tail fiber modifying protein. *The EMBO journal*, 4(9): 2343–6.

Rifat, D., Wright, N.T., Varney, K.M., Weber, D.J. & Black, L.W. 2008. Restriction Endonuclease Inhibitor IPI* of Bacteriophage T4: A Novel Structure for a Dedicated Target. *Journal of Molecular Biology*, 375(3): 720–734.

Riley, P.A. 2006. Phages: their role in bacterial pathogenesis and biotechnology. *Journal of Clinical Pathology*, 59(9): 1003–1004.

Ripp, S. & Miller, R. V. 1998. Dynamics of the pseudolysogenic response in slowly growing cells of Pseudomonas aeruginosa. *Microbiology*, 144(8): 2225–2232.

Roach, D.R., Khatibi, P.A., Bischoff, K.M., Hughes, S.R., Donovan, D.M., Ugarte, D.D.L.T., English, B., Jensen, K., Connolly, C., Skinner, K., Leathers, T., Schell, D., Dowe, N., Ibsen, K., Riley, C., Ruth, M., Lumpkin, R., Beckner, M., Ivey, M., Phister, T., Bischoff, K., Liu, S., Leathers, T., Worthington, R., Rich, J., Limayem, A., Hanning, I., Muthaiyan, A., Illeghems, K., Kim, J.-W., Crandall, P., O, C., Chang, I.-S., Kim, B.-H., Shin, P.-K., Lee, Y.-L., Lushia, W., Heist, P., Narendranath, N., Makanjuola, D., Tymon, A., Springham, D., Narendranath, N., Hynes, S., Thomas, K., Ingledew, W., Schnurer, J., Magnusson, J., Lindgren, S., Dobrogosz, W., Sjogren, J., Magnusson, J., Broberg, A., Schnurer, J., Kenne, L., Magnusson, J., Strom, K., Sogren, S., Schnurer, J., Bischoff, K., Skinner-Nemec, K., Leathers, T., McChesney, D., Bernhardt, T., Wang, I., Struck, D., Young, R., Loeffler, J., Nelson, D., Fischetti, V., Nelson, D., Loomis, L., Fischetti, V., Schuch, R., Nelson, D., Fischetti, V., Schmelcher, M., Powell, A., Becker, S., Camp, M., Donovan, D., Nelson, D., Schmelcher, M., Rodriguez-Rubio, L., Klumpp, J., Pritchard, D., Dong, S., Donovan, D., Schleifer, K., Kandler, O., Fischetti, V., Turner, M., Hafner, L., Walsh, T., Giffard, P., Ribelles, P., Rodriguez, I., Suarez, J., Yokoi, K., Shinohara, M., Kawahigashi, N., Nakagawa, K., Kawasaki, K., Nakamura, S., Taketo, A., Kodaira, K., Pritchard, D., Dong, S., Kirk, M., Cartee, R., Baker, J., Donovan, D., Foster-Frey, J., Wang, S., Kong, J., Zhang, X., Bischoff, K., Poole, T., Beier, R., Spratt, B., Gerboux, V., Villa, A., Monany, C., Bertrand, A., Delfini, C., Cersosimo, M., Prete, V. Del, Strano, M., Gaetano, G., Pagliara, A., Ambro, S., Blattel, V., Wirth, K., Claus, H., Schlott, B., Pfeiffer, P., Konig, H., Hoopes, J., Stark, C., Kim, H., Sussman, D., Donovan, D., Nelson, D., Graham, S., Coote, P., Eugster, M., Haug, M., Huwiler, S., Loessner, M., Russell, I., Villion, M., Moineau, S., Henrich, B., Binishofer, B., Blasi, U., Oki, M., Kakikawa, M., Yamada, K., Taketo, A., Kodaira, K., Kakikawa, M., Yokoi, K., Kimoto, H., Nakano, M., Kawasaki, K., Taketo, A., Kodaira, K., Kashige, N., Nakashima, Y., Miake, F.,

Watanabe, K., Deutsch, S., Guezenec, S., Piot, M., Foster, S., Lortal, S., Becker, S., Foster-Frey, J., Stodola, A., Anacker, D., Donovan, D., Becker, S., Foster-Frey, J., Donovan, D., Bischoff, K., Liu, S., Hughes, S. & Rich, J. 2013. Bacteriophage-encoded lytic enzymes control growth of contaminating Lactobacillus found in fuel ethanol fermentations. *Biotechnology for Biofuels*, 6(1): 20.

Roessner, C.A. & Ihler, G.M. 1986. Formation of transmembrane channels in liposomes during injection of lambda DNA. *The Journal of biological chemistry*, 261(1): 386–90.

Romantschuk, M. & Bamford, D.H. 1985. Function of pili in bacteriophage phi 6 penetration. *The Journal of general virology*, 66(11): 2461–9.

Romero, P., López, R. & García, E. 2004. Genomic organization and molecular analysis of the inducible prophage EJ-1, a mosaic myovirus from an atypical pneumococcus. *Virology*, 322(2): 239–52.

Ronca, S., Ramond, J.-B., Jones, B.E., Seely, M. & Cowan, D.A. 2015. Namib Desert dune/interdune transects exhibit habitat-specific edaphic bacterial communities. *Frontiers in microbiology*, 6: 845.

Rosario, K., Nilsson, C., Lim, Y.W., Ruan, Y. & Breitbart, M. 2009. Metagenomic analysis of viruses in reclaimed water. *Environmental Microbiology*, 11(11): 2806–2820.

Rose, K., Mason, J.O. & Lathe, R. 2002. Hybridization parameters revisited: Solutions containing SDS. *BioTechniques*, 33(1): 54–58.

Rossmann, M.G., Mesyanzhinov, V. V., Arisaka, F. & Leiman, P.G. 2004. The bacteriophage T4 DNA injection machine. *Current Opinion in Structural Biology*, 14(2): 171–180.

Roux, S., Tournayre, J., Mahul, A., Debroas, D. & Enault, F. 2014. Metavir 2: new tools for viral metagenome comparison and assembled virome analysis. *BMC bioinformatics*, 15(1): 76.

Rydman, P.S., Caldentey, J., Butcher, S.J., Fuller, S.D., Rutten, T. & Bamford, D.H. 1999. Bacteriophage PRD1 contains a labile receptor-binding structure at each vertex. *Journal of molecular biology*, 291(3): 575–87.

Sabban, S. 2011. *Development of an in vitro model system for studyng the interaction of Equus caballus IgE with its high-affinity Fc receptor*. The University of Sheffield.

Saha, S., Haggard-Ljungquist, E. & Nordstrom, K. 1989. Activation of prophage P4 by the P2 Cox protein and the sites of action of the Cox protein on the two phage genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 86(11): 3973–3977.

Saha, S., Haggård-Ljungquist, E. & Nordström, K. 1987. The cox protein of bacteriophage P2 inhibits the formation of the repressor protein and autoregulates the early operon. *The EMBO journal*, 6(10): 3191–3199.

Saitou, N. & Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*, 4(4): 406–25.

Sakkhachornphop, S., Jiranusornkul, S., Kodchakorn, K., Nangola, S., Sirisanthana, T. & Tayapiwatana, C. 2009. Designed zinc finger protein interacting with the HIV-1 integrase recognition sequence at 2-LTR-circle junctions. *Protein Science*, 18(11): 2219–2230.

Sambrook, J., Fritsch, E. & Maniatis, F. 1989. Molecular Cloning: a Laboratory Manual.

Sambrook, J., Fritsch, E.F. & Maniatis, T. 1989. Molecular cloning: a laboratory manual. *Cold Spring Harbor Laboratory Press*: 626.

Samson, J.E., Magadán, A.H., Sabri, M. & Moineau, S. 2013. Revenge of the phages: defeating bacterial defences. *Nature reviews. Microbiology*, 11(10): 675–87.

Samsygina, G.A. & Boni, E.G. 1984. [Bacteriophages and phage therapy in pediatric practice]. *Pediatriia*, 4(4): 67–70.

Sandegren, L. & Sjöberg, B.M. 2007. Self-splicing of the bacteriophage T4 group I introns requires efficient translation of the pre-mRNA in vivo and correlates with the growth state of the infected bacterium. *Journal of Bacteriology*, 189(3): 980–990.

Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, a R., Fiddes, C. a, Hutchison, C. a,

Slocombe, P.M. & Smith, M. 1977. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature*, 265(5596): 687–695.

Sanger, F., Coulson, A.R., Hong, G.F., Hill, D.F. & Petersen, G.B. 1982. Nucleotide sequence of bacteriophage lambda DNA. *Journal of Molecular Biology*, 162(4): 729–773.

Sano, E., Carlson, S., Wegley, L. & Rohwer, F. 2004. Movement of viruses between biomes. *Applied and Environmental Microbiology*, 70(10): 5842–5846.

Santos, F., Yarza, P., Parro, V., Briones, C. & Antón, J. 2010. The metavirome of a hypersaline environment. *Environmental microbiology*, 12(11): 2965–76.

Sato, K. & Campbell, A. 1970. Specialized transduction of galactose by lambda phage from a deletion lysogen. *Virology*, 41(3): 474–487.

Schattner, P., Brooks, A.N. & Lowe, T.M. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research*, 33(SUPPL. 2): W686–9.

Schoenfeld, T., Patterson, M., Richardson, P.M., Wommack, K.E., Young, M. & Mead, D. 2008. Assembly of viral metagenomes from Yellowstone hot springs. *Applied and Environmental Microbiology*, 74(13): 4164–4174.

Schwartz, M. 1975. Reversible interaction between coliphage lambda and its receptor protein. *Journal of molecular biology*, 99(1): 185–201.

Seely, M.. & Henschel, J.R. 2000. Developing fog harvesting in Namibia. *4th Biennial Congress of the African Division of the International Association of Hydraulic Research, Windhoek, Namibia, Newsletter of the Namibia Scientific Society*, (41): 30– 42.

Serwer, P., Hayes, S.J., Thomas, J. a & Hardies, S.C. 2007. Propagating the missing bacteriophages: a large bacteriophage in a new class. *Virology journal*, 4: 21.

Shan, J., Patel, K. V., Hickenbotham, P.T., Nale, J.Y., Hargreaves, K.R. & Clokie, M.R.J. 2012. Prophage carriage and diversity within clinically relevant strains of Clostridium difficile. *Applied*

and Environmental Microbiology, 78(17): 6027–6034.

Shen, P.S., Domek, M.J., Sanz-Garcia, E., Makaju, a., Taylor, R.M., Hoggan, R., Culumber, M.D., Oberg, C.J., Breakwell, D.P., Prince, J.T. & Belnap, D.M. 2012. Sequence and Structural Characterization of Great Salt Lake Bacteriophage CW02, a Member of the T7-Like Supergroup. *Journal of Virology*, 86(15): 7907–7917.

Short, J.M., Fernandez, J.M., Sorge, J.A. & Huse, W.D. 1988. Lambda ZAP: a bacteriophage lambda expression vector with in vivo excision properties. *Nucleic Acids Research*, 16(15): 7583–7600.

Shotland, Y., Koby, S., Teff, D., Mansur, N., Oren, D.A., Tatematsu, K., Tomoyasu, T., Kessel, M., Bukau, B., Ogura, T. & Oppenheim, A.B. 1997. Proteolysis of the phage lambda CII regulatory protein by FtsH (HflB) of Escherichia coli. *Molecular microbiology*, 24(6): 1303–1310.

Sidhu, S. 2000. Phage display in pharmaceutical biotechnology. *Current Opinion in Biotechnology*, 11(6): 610–616.

Silahtaroglu, A., Pfundheller, H., Koshkin, A., Tommerup, N. & Kauppinen, S. 2004. LNA-modified oligonucleotides are highly efficient as FISH probes. *Cytogenetic and genome research*, 107(1-2): 32–7.

Simon, C. & Daniel, R. 2011. Metagenomic analyses: Past and future trends. *Applied and Environmental Microbiology*, 77(4): 1153–1161.

Simon, L.D. & Anderson, T.F. 1967a. The infection of Escherichia coli by T2 and T4 bacteriophages as seen in the electron microscope I. Attachment and penetration. *Virology*, 32(2): 279–297.

Simon, L.D. & Anderson, T.F. 1967b. The infection of Escherichia coli by T2 and T4 bacteriophages as seen in the electron microscope II. Structure and function of the baseplate. *Virology*, 32(2): 298–305.

Sleator, R.D., Shortall, C. & Hill, C. 2008. Metagenomics. *Letters in Applied Microbiology*, 47(5): 361–366.

Smith, G.P. 1985. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*, 228(4705): 1315–1317.

Smolina, I. V., Cherny, D.I., Nietupski, R.M., Beals, T., Smith, J.H., Lane, D.J., Broude, N.E. & Demidov, V. V. 2005. High-density fluorescently labeled rolling-circle amplicons for DNA diagnostics. *Analytical Biochemistry*, 347(1): 152–155.

Sorek, R., Kunin, V. & Hugenholtz, P. 2008. CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nature reviews. Microbiology*, 6(3): 181–186.

Sorge, J.A. 1988. Bacteriophage lambda cloning vectors. *Biotechnology*, 10(0740-7378): 43–60.

Stent, G.S. 1963. Molecular biology of bacterial viruses. *Molecular biology of bacterial viruses.*

Stern, A., Mick, E., Tirosh, I., Sagy, O. & Sorek, R. 2012. CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome. *Genome research*, 22(10): 1985–94.

Steward, G.F., Culley, A.I., Mueller, J.A., Wood-Charlson, E.M., Belcaid, M. & Poisson, G. 2013. Are we missing half of the viruses in the ocean? *The International Society for Microbial Ecology journal*, 7(3): 672–9.

Stomeo, F., Valverde, A., Pointing, S.B., McKay, C.P., Warren-Rhodes, K.A., Tuffin, M.I., Seely, M. & Cowan, D.A. 2013. Hypolithic and soil microbial community assembly along an aridity gradient in the Namib Desert. *Extremophiles*, 17(2): 329–337.

Strandberg, L. & Enfors, S.O. 1991. Factors influencing inclusion body formation in the production of a fused protein in Escherichia coli. *Applied and Environmental Microbiology*, 57(6): 1669–1674.

Streefkerk, J.G. 1972. Inhibition of erthrocyte pseudoperoxidase activity by treatment with hydrogen peroxidase following methanol. *Nature*, (330): 80.

Streisinger, G., Emrich, J. & Stahl, M.M. 1967. Chromosome structure in phage T4, iii. Terminal redundancy and length determination. *Proceedings of the National Academy of Sciences of the United States of America*, 57(2): 292–5.

Sulakvelidze, A. 2011. The challenges of bacteriophage therapy. *Industrial Pharmacy*, (31): 14–18.

Sulakvelidze, A., Alavidze, Z. & Morris, J.G.J. 2001. Bacteriophage Therapy. *Antimicrobial Agents and Chemotherapy*, 45(3): 649–659.

Sullivan, M.B., Coleman, M.L., Quinlivan, V., Rosenkrantz, J.E., DeFrancesco, A.S., Tan, G., Fu, R., Lee, J.A., Waterbury, J.B., Bielawski, J.P. & Chisholm, S.W. 2008. Portal protein diversity and phage ecology. *Environmental Microbiology*, 10(10): 2810–2823.

Sullivan, M.J., Petty, N.K. & Beatson, S.A. 2011. Easyfig: a genome comparison visualizer. *Bioinformatics*, 27(7): 1009–1010.

Susskind, M.M. & Botstein, D. 1978. Molecular genetics of bacteriophage P22. *Microbiological reviews*, 42(2): 385–413.

Suttle, C.A. & Chan, A.M. 1994. Dynamics and distribution of cyanophages and their effect on marine Synechococcus spp. *Applied and Environmental Microbiology*, 60(9): 3167–3174.

Svenningsen, S. Lo & Semsey, S. 2014. Commitment to lysogeny is preceded by a prolonged period of sensitivity to the late lytic regulator Q in bacteriophage λ. *Journal of Bacteriology*, 196(20): 3582–3588.

Swan, B.K., Ehrhardt, C.J., Reifel, K.M., Moreno, L.I. & Valentine, D.L. 2010. Archaeal and bacterial communities respond differently to environmental gradients in anoxic sediments of a California hypersaline lake, the Salton Sea. *Applied and environmental microbiology*, 76(3): 757–68.

Swarts, D.C., Mosterd, C., van Passel, M.W.J. & Brouns, S.J.J. 2012. CRISPR interference directs strand specific spacer acquisition. *PLoS ONE*, 7(4): 1–7.

Szczepankowska, A. 2012. Chapter 8 – Role of CRISPR/cas System in the Development of Bacteriophage Resistance. In *Advances in Virus Research*. 289–338.

Tabor, S. 2001. Expression using the T7 RNA polymerase/promoter system. *Current protocols in molecular biology*, Chapter 16: Unit16.2.

Tadmor, A.D., Ottesen, E.A., Leadbetter, J.R. & Phillips, R. 2011. Probing individual environmental bacteria for viruses by using microfluidic digital PCR. *Science*, 333(6038): 58–62.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30(12): 2725–2729.

Tanji, Y., Furukawa, C., Na, S.H., Hijikata, T., Miyanaga, K. & Unno, H. 2004. Escherichia coli detection by GFP-labeled lysozyme-inactivated T4 bacteriophage. *Journal of Biotechnology*, 114(1-2): 11–20.

Tarahovsky, Y.S., Khusainov, A.A., Deev, A.A. & Kim, Y. V. 1991. Membrane fusion during infection of Escherichia coli cells by phage T4. *FEBS letters*, 289(1): 18–22.

Teeling, H., Waldmann, J., Lombardot, T., Bauer, M., Glöckner, F., Amann, R., Ludwig, W., Schleifer, K., Rondon, M., August, P., Bettermann, A., Brady, S., Grossman, T., Liles, M., Loiacono, K., Lynch, B., MacNeil, I., Minor, C., Tiong, C., Gilman, M., Osburne, M., Clardy, J., Handelsman, J., Goodman, R., Beja, O., Aravind, L., Koonin, E., Suzuki, M., Hadd, A., Nguyen, L., Jovanovich, S., Gates, C., Feldman, R., Spudich, J., Spudich, E., DeLong, E., Beja, O., Spudich, E., Spudich, J., Leclerc, M., DeLong, E., Torre, J. de la, Christianson, L., Beja, O., Suzuki, M., Karl, D., Heidelberg, J., DeLong, E., Sabehi, G., Massana, R., Bielawski, J., Rosenberg, M., Delong, E., Beja, O., Tyson, G., Chapman, J., Hugenholtz, P., Allen, E., Ram, R., Richardson, P., Solovyev, V., Rubin, E., Rokhsar, D., Banfield, J., Teeling, H., Meyerdierks, A., Bauer, M., Amann, R., Glockner, F., Karlin, S., Ladunga, I., Blaisdell, B., Karlin, S., Burge, C., Karlin, S., Karlin, S., Campbell, A., Mrazek, J., Nakashima, H., Ota, M., Nishikawa, K., Ooi, T., Gentles, A., Karlin, S., Abe, T., Kanaya, S., Kinouchi, M., Ichiba, Y., Kozuki, T., Ikemura, T., Deschavanne, P., Giron, A., Vilain, J., Fagot, G., Fertil, B., Goldman, N., Sandberg, R., Winberg, G., Branden, C., Kaske, A., Ernberg, I., Coster, J., Pride, D., Meinersmann, R., Wassenaar, T., Blaser, M., Schbath, S., Prum, B., Turckheim, E. de & Schbath, S. 2004. TETRA: a web-service and a stand-alone program for the analysis and comparison of tetranucleotide usage patterns in DNA sequences. *BMC Bioinformatics*, 5(1): 163.

Terzaghi, B.E., Terzaghi, E. & Coombs, D. 1979. The role of the collar/whisker complex in bacteriophage T4D tail fiber attachment. *Journal of Molecular Biology*, 127(1): 1–14.

Thingstad, T.F. 2000. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnology and Oceanography*, 45(6): 1320–1328.

Thomas, A. 2013. Flow cytometry. : 1–3. http://www.serpins.com/prod02.htm 1 January 2015.

Thurber, R.V. 2009. Current insights into phage biodiversity and biogeography. *Current Opinion in Microbiology*, 12(5): 582–587.

Thurnheer, T., Gmür, R. & Guggenheim, B. 2004. Multiplex FISH analysis of a six-species bacterial biofilm. *Journal of Microbiological Methods*, 56(1): 37–47.

Tillier, E.R. & Collins, R.A. 2000. The contributions of replication orientation, gene direction, and signal sequences to base-composition asymmetries in bacterial genomes. *Journal of molecular evolution*, 50(3): 249–57.

Todar, K. 2001. The Microbial World. : 1–7. http://www.textbookofbacteriology.net/themicrobialworld/Phage.html 15 April 2013.

Tolonen, N., Doglio, L., Schleich, S. & Krijnse Locker, J. 2001. Vaccinia virus DNA replication occurs in endoplasmic reticulum-enclosed cytoplasmic mini-nuclei. *Molecular biology of the cell*, 12(7): 2031–2046.

Traganos, F. 1984. *Flow Cytometry: Principles and Applications. I*. Humana Press.

Tufariello, J.M., Malek, A.A., Vilcheze, C., Cole, L.E., Ratner, H.K., Gonzalez, P.A., Jain, P., Hatfull, G.F., Larsen, M.H. & Jacobs, W.R. 2014. Enhanced Specialized Transduction Using Recombineering in Mycobacterium tuberculosis. *mBio*, 5(3): e01179–14–e01179–14.

Twort, F.W. 1915. An Investigation on the Nature of Ultra-Microscopic Viruses. *The Lancet*, 186(4814): 1241–1243.

Tyson, G.W., Lo, I., Baker, B.J., Allen, E.E., Hugenholtz, P. & Banfield, J.F. 2005. Genome-directed isolation of the key nitrogen fixer Leptospirillum ferrodiazotrophum sp. nov. from an acidophilic microbial community. *Applied and environmental microbiology*, 71(10): 6319–24.

Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S. & Gregory, P.D. 2010. Genome editing with engineered zinc finger nucleases. *Nature Reviews Genetics*, 11(9): 636–646.

Valpuesta, J.M. & Carrascosa, J.L. 1994. Structure of viral connectors and their function in bacteriophage assembly and DNA packaging. *Quarterly reviews of biophysics*, 27(2): 107–155.

Vareschi, E. & Jacobs, J. 1985. The ecology of Lake Nakuru. *Oecologia*, 65(3): 412–424.

Vartoukian, S.R., Palmer, R.M. & Wade, W.G. 2010. Strategies for culture of 'unculturable' bacteria. *FEMS Microbiology Letters*, 309(1): 1–7.

Vasu, K. & Nagaraja, V. 2013. Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiology and molecular biology reviews*, 77(1): 53–72.

Vieira, J. & Messing, J. 1987. *Production of single stranded plasmid DNA*. Elsevier.

Vikram, S., Guerrero, L.D., Makhalanyane, T.P., Le, P.T., Seely, M. & Cowan, D.A. 2015. Metagenomic analysis provides insights into functional capacity in a hyperarid desert soil niche community. *Environmental Microbiology*, 18(6): 1875–88.

Viles, H.A. & Goudie, A.S. 2013. Weathering in the central Namib Desert, Namibia: Controls, processes and implications. *Journal of Arid Environments*, 93: 20–29.

Villarroel, J., Kleinheinz, K., Jurtz, V., Zschach, H., Lund, O., Nielsen, M. & Larsen, M. 2016. HostPhinder: A Phage Host Prediction Tool. *Viruses*, 8(5): 116.

Ward, S., Luftig, R.B., Wilson, J.H., Eddleman, H., Lyle, H. & Wood, W.B. 1970. Assembly of bacteriophage T4 tail fibers: II. Isolation and characterization of tail fiber precursors. *Journal of Molecular Biology*, 54(1): 15–31.

Waterbury, J.B. & Valois, F.W. 1993. Resistance to co-occurring phages enables marine

Synechococcus communities to coexist with cyanophages abundant in seawater. *Applied and Environmental Microbiology*, 59(10): 3393–3399.

Waterbury, J.B., Watson, S.W., Valois, F.W. & Franks, D.G. 1986. Biological and ecological characterisation of the marine unicellular cyanobacterium Synechococcus. *Photosynthetic Picoplankton*, 214: 71–120.

Watson, B. & Eveland, W. 1965. The application of the phage-fluorescent antiphage staining system in the specific identification of Listeria monocytogenes. I. Species specificity and immunofluorescent sensitivity of Listeria monocytogenes phage observed in smear preparations. *The Journal of infectious diseases*, 115 (4): 363–369.

Webster, R. 1996. Biology of filamentous bacteriophage. In BK Kay, ed. *Phage Display of Peptides and Proteins*. Academic Press: 1–20.

Wegrzyn, G. & Wegrzyn, A. 2005. Genetic switches during bacteriophage lambda development. *Progress in nucleic acid research and molecular biology*, 79: 1–48.

Wei, Y., Terns, R.M., Terns, M.P., Terns, M.P. & Terns, M.P. 2015. Cas9 function and host genome sampling in type II-A CRISPR–cas adaptation. , 29(4): 356–361.

Weisberg, R.A. & Landy, A. 1983. Site-specific Recombination in Phage Lambda. In R. Hendrix, J. Roberts, F. Stahl, & R. Weisberg, eds. *Lambda II*. New York: Cold Spring Harbor Press: 211–250.

Werts, C., Michel, V., Hofnung, M. & Charbit, A. 1994. Adsorption of bacteriophage lambda on the LamB protein of Escherichia coli K-12: Point mutations in gene J of lambda responsible for extended host range. *Journal of Bacteriology*, 176(4): 941–947.

Wilhelm, S.W. & Suttle, C. a. 1999. Viruses and Nutrient Cycles in the Sea. *BioScience*, 49(10): 781.

von Willert, D.J. 1992. *Life Strategies of Succulents in Deserts: With Special Reference to the Namib Desert*. CUP Archive.

Williamson, S.J., Rusch, D.B., Yooseph, S., Halpern, A.L., Heidelberg, K.B., Glass, J.I., Pfannkoch, C.A., Fadrosh, D., Miller, C.S., Sutton, G., Frazler, M. & Venter, J.C. 2008. The sorcerer II global ocean sampling expedition: Metagenomic characterization of viruses within aquatic microbial samples. *PLoS ONE*, 3(1): e1456.

Wilson, W.H., Carr, N.G. & Mann, N.H. 1996. the Effect of Phosphate Status on the Kinetics of Cyanophage Infection in the Oceanic Cyanobacterium Synechococcus Sp. Wh78031. *Journal of Phycology*, 32(4): 506–516.

Wommack, K.E. & Colwell, R.R. 2000. Virioplankton: viruses in aquatic ecosystems. *Microbiology and molecular biology reviews : MMBR*, 64(1): 69–114.

Wommack, K.E., Ravel, J., Hill, R.T. & Colwell, R.R. 1999. Hybridization analysis of Chesapeake Bay virioplankton. *Applied and Environmental Microbiology*, 65(1): 241–250.

Wurch, L., Giannone, R.J., Belisle, B.S., Swift, C., Utturkar, S., Hettich, R.L., Reysenbach, A.-L. & Podar, M. 2016. Genomics-informed isolation and characterization of a symbiotic Nanoarchaeota system from a terrestrial geothermal environment. *Nature Communications*, 7: 12115.

Xu, J., Hendrix, R.W. & Duda, R.L. 2004. Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. *Molecular Cell*, 16(1): 11–21.

Yanagida, M. & Ahmad-Zadeh, C. 1970. Determination of gene product positions in bacteriophage T4 by specific antibody association. *Journal of Molecular Biology*, 51(2): 411–421.

Yarmolinsky, M.B. 2004. Bacteriophage P1 in retrospect and in prospect. *Journal of bacteriology*, 186(21): 7025–8.

Yassin, A. & Mankin, A.S. 2007. Potential new antibiotic sites in the ribosome revealed by deleterious mutations in RNA of the large ribosomal subunit. *Journal of Biological Chemistry*, 282(33): 24329–24342.

Yokouchi, H., Fukuoka, Y., Mukoyama, D., Calugay, R., Takeyama, H. & Matsunaga, T. 2006. Whole-metagenome amplification of a microbial community associated with scleractinian coral by

multiple displacement amplification using ??29 polymerase. *Environmental Microbiology*, 8(7): 1155–1163.

Yoshida, M., Takaki, Y., Eitoku, M., Nunoura, T. & Takai, K. 2013. Metagenomic Analysis of Viral Communities in (Hado)Pelagic Sediments. *PLoS ONE*, 8(2): e57271.

Yu, F., Tanaka, Y., Yamashita, K., Suzuki, T., Nakamura, a., Hirano, N., Suzuki, T., Yao, M. & Tanaka, I. 2011. Molecular basis of dihydrouridine formation on tRNA. *Proceedings of the National Academy of Sciences*, 108(49): 19593–19598.

Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J. & Wishart, D.S. 2011. PHAST: A Fast Phage Search Tool. *Nucleic Acids Research*, 39(SUPPL. 2): W347–52.

Zimmer, C. & Wahnert, U. 1986. Nonintercalating DNA-binding ligands: Specificity of the interaction and their use as tools in biophysical, biochemical and biological investigations of the genetic material. *Progress in Biophysics and Molecular Biology*, 47(1): 31–112.

van Zyl, L.J., Sunda, F., Taylor, M.P., Cowan, D.A. & Trindade, M.I. 2015. Identification and characterization of a novel Geobacillus thermoglucosidasius bacteriophage, GVE3. *Archives of Virology*, 160(9): 2269–2282.

Van Zyl, L.J., Taylor, M.P., Eley, K., Tuffin, M. & Cowan, D.A. 2014. Engineering pyruvate decarboxylase-mediated ethanol production in the thermophilic host Geobacillus thermoglucosidasius. *Applied Microbiology and Biotechnology*, 98(3): 1247–1259.

van Zyl, L.J., Taylor, M.P. & Trindade, M. 2016. Engineering resistance to phage GVE3 in Geobacillus thermoglucosidasius. *Applied microbiology and biotechnology*, 100(4): 1833–41.